ANALYSIS OF VARIANCE

N GIRI

SUBSIDISED
LOW COST
L C
S E
EDITIONS
NBT INDIA

ANALYSIS
of
VARIANCE

N.GIRI

is Professor of Mathematics at the University of Montreal. He has taught statistics at the Indian Statistical Institute, Forschungsinstitut für Mathematik (Zurich), the Indian Institute of Technology, Kanpur, Cornell University and the University of Arizona. He received his B.Sc. in Mathematics from Midnapur College (India) in 1951, M.Sc. in Statistics from Calcutta University in 1953 and Ph.D. in Statistics from Stanford University in 1961, working with Professor Charles Stein.

The author's research interests centre on multivariate analysis and design of experiments. He has published more than 46 original research papers on these and related topics in Annals of Mathematical Statistics, Archiv der Mathematik, Journal of American Statistical Association, Sankhya, Journal of Agricultural Research Statistics, Annals of Statistical Mathematics, Biometrics, etc. He has published four books: Introduction to Statistics and Probability, parts I and II, Invariance and Minimax Statistical Tests, Multivariate Statistical Inference and Design and Analysis of Experiments (co-author Dr.M.N.Das).

He is a fellow of the Royal Statistical Society (UK), an honorary fellow of the American Statistical Association and of the Institute of Mathematical Statistics. He is also an elected member of the International Statistical Institute.

ANALYSIS
of
VARIANCE

B11371

N. GIRI

Department of Mathematics & Statistics
University of Montreal, Montreal, Quebec H3C 3J7

CENTRAL LIBRARY
St. Xavier's College, Kolkata

Central Library
B11371
519.538 GIR(ANA)
St. Xavier's College

South Asian Publishers – New Delhi

Central Library
B11371
519.538 GIR(ANA)

Copyright © South Asian Publishers, 1986

No part of this book may be reproduced in any form
and by any means without the written permission
of the publishers

This book has been subsidized by the Government
of India, through the National Book Trust, India,
for the benefit of students

Price : Rs. 36.00

NBT Code No. 45-233/1982

ISBN 81-7003-063-3

Published by South Asian Publishers Pvt. Ltd.,
36, Netaji Subhash Marg, Daryaganj,
New Delhi 110002 and printed at Chaman Offset
Printers, 1626 Suiwalan, Daryaganj,
New Delhi 110002. Printed in India.

TO

NILIMA

NABANITA

& NANDAN

PREFACE

This book deals with an up-to-date presentation of both theoretical and applied aspects of analysis of variance in univariate and multivariate cases. Various designs are explained with live data from different applied fields. The materials presented here are developed from the lecture notes of a one semester course at the graduate level on analysis of variance which I presented several times at the University of Montreal, and in parts at the Indian Statistical Institute, Calcutta, Forschungsinstitut der Mathematik, the University of Arizona and Dibrugarh University, Assam, India. Each chapter contains numerous related problems, complete references and worked out examples. The exercises of chapters 6 and 7 on multivariate analysis of variance also contain analogous result for the complex multivariate normal populations.

Preparation and revision of the manuscript would not have been an easy task without the help of Dr. M.N.Das who helped me with great care and diligence. This book was written with the direct and indirect help of many people. I owe a great debt to my parents and brothers and sisters for their help. My wife Nilima, daughter Nabanita and son Nandan have been very helpful during the preparation of the book. I gratefully acknowledge their assistance.

Montreal
March 1985

N.GIRI

CONTENTS

Chapter 1 SOME MATHEMATICAL CONCEPTS 1
- 1.1 Matrix Algebra, 1
- 1.2 Groups, 21
- Exercise 1, 25
- References, 27

Chapter 2 BASIC STATISTICAL CONCEPTS 29
- 2.1 Noncentral Chisquare Distribution, 29
- 2.2 Noncentral F, Noncentral Beta and Noncentral t-Distributions, 32
- 2.3 Multivariate Normal Distribution, Quadratic Form and Cochran's Theorem, 34
- 2.4 Invariance in Statistical Testing, 54
- Exercise 2, 81
- References, 85

Chapter 3 UNIVARIATE GENERAL LINEAR HYPOTHESIS AND ANALYSIS OF VARIANCE 87
- 3.1 General Linear Hypothesis, 87
- 3.2 Regression Model of the General Linear Hypothesis, 93
- 3.3 Percent Points and Confidence Interval, 118
- 3.4 Tests of Hypotheses in the Regression Model, 121
- 3.5 One-way Classifications with Linear Models of Full Rank, 128
- 3.6 Experimental Design Models, 141
- 3.7 Two-way Classification, 148
- 3.8 Three-way Classification with Interactions, 158
- Exercise 3, 162
- References, 171

Chapter 4 ANALYSIS OF COVARIANCE OPTIMALITY OF DESIGNS 173
- 4.1 Analysis of Covariance, 174
- 4.2 Efficiency, Optimality of Designs, 187
- Exercise 4, 191
- References, 194

Chapter 5 EXPERIMENTAL DESIGNS AND THEIR ANALYSIS 196
- 5.1 Completely Randomized Design, 197
- 5.2 Randomized Block Design, 199
- 5.3 Latin Squares Design, 224
- 5.4 Balanced Incomplete Block Designs, 244
- 5.5 Partially Balanced Incomplete Block (PBIB) Designs, 273
- Exercise 5, 291
- References, 297

Chapter 6 ANALYSIS OF VARIANCE II, FACTORIAL EXPERIMENTS 304

6.1 2^n-Factorial Experiments, 304
6.2 General Factorial Experiments, 307
6.3 Confounding, 310
6.4 Nested Designs (Fixed Effect Model), 314
6.5 Random-effect Model, 318
6.6 Nested Design with Random-effect Model, 322
6.7 Crossed and Nested Designs (Mixed Effect Model), 325
6.8 Split-plot Designs, 328
6.9 Lattice Designs, 331
Exercise 6, 335
References, 337

Chapter 7 MULTIVARIATE GENERAL LINEAR HYPOTHESIS WITH APPLICATIONS 339

7.1 General Linear Hypothesis (Abstract Formulation), 339
7.2 The Distribution of $R_1,\ldots,R_p$ under H_o, 348
7.3 Distribution of Test Statistics under H_o, 351
7.4 Optimum Properties of Tests of General Linear Hypothesis, 353
7.5 Multivariate Regression Model, 354
7.6 Applications of Multivariate General Linear Hypothesis, 361
7.7 Multivariate Analysis of Covariance, 379
7.8 Case of Multinomial and Complex Multivariate Normal Distributions, 380
7.9 Statistical Tables on Wilks' Criterion $U=U_{p,r,N-s}$, 381
Exercise 7, 393
References, 394

Chapter 8 NONPARAMETRIC ANALYSIS OF VARIANCE 399

8.1 One-sample Location Parameter (Estimation), 399
8.2 Two-sample Location Parameters (Estimation and Testing), 402
8.3 One-way classification (Location Model), 407
8.4 Experiments with Two or More Factors, 412
Exercise 8, 419
References, 423

Author Index 426

Subject Index 428

CHAPTER 1. SOME MATHEMATICAL CONCEPTS

The univariate analysis of variance, as commonly understood and practised today in applied sciences, has been developed chiefly by Professor R.A. Fisher, who first introduced the terms analysis of variance and variance components in statistics. It is a statistical technique for analyzing a set of observations depending on several kinds of effects, not necessarily independent and operating simultaneously, to decide which kinds of effects are of importance and to estimate them. Essentially it involves decomposing the total variation present in the set of observations into several meaningful variance components, representing different effects, which generate the observations and making suitable comparisons among these components.

If each observation is a vector quantity depending on several characteristics, we get a multivariate analogue of this technique, and this is called the multivariate analysis of variance. In this chapter we shall present some basic mathematical results of matrix algebra and group theory which are essential to understand the material presented in this book.

1.1. MATRIX ALGEBRA

A real matrix A of dimension $p \times q$ is an ordered rectangular array of real elements a_{ij}, $i = 1,\dots,p$, $j = 1,\dots,q$, and is written as

$$A = (a_{ij}) = \begin{pmatrix} a_{11},\dots,a_{1q} \\ \dots,\dots,\dots \\ a_{p1},\dots,a_{pq} \end{pmatrix}.$$

The transpose of A is a $q \times p$ matrix $A' = (a_{ji})$. A matrix A with $p = q$ is called a square matrix. A square matrix A is said to be symmetric if $A = A'$ and skew symmetric if $A = -A'$. Obviously the diagonal elements of a skew symmetric matrix are zeros.

Diagonal matrix: A square matrix whose off-diagonal elements are zeros is called a diagonal matrix.

A diagonal matrix whose diagonal elements are unity is called an identity matrix and is denoted by I.

Triangular matrix: A square matrix $A = (a_{ij})$ with $a_{ij} = 0$, $j < i$ is an upper triangular matrix. If $a_{ij} = 0$ for $j > i$, then A is called a lower triangular matrix.

Orthogonal matrix: A square matrix A is orthogonal if $AA' = A'A = I$.

Nonsingular matrix: A square matrix A whose determinant, denoted by $|A|$, is not equal to zero, is called a nonsingular matrix. If $|A| = 0$, then $|A|$ is called a singular matrix. The rows and columns of a nonsingular matrix are linearly independent. The sum of any two nonsingular matrices is not necessarily nonsingular.

Theorem 1.1. Let A and B be two matrices of dimensions $p \times q$ and $q \times p$ respectively. Then $|I_p + AB| = |I_q + BA|$, where I_p and I_q are identity matrices of dimensions $p \times p$ and $q \times q$ respectively.

Proof.

$$\begin{vmatrix} -I_p, A \\ B, I_q \end{vmatrix} = \begin{vmatrix} -I_q, B \\ A, I_p \end{vmatrix}$$

But

$$\begin{vmatrix} -I_p, A \\ B, I_q \end{vmatrix} = \begin{vmatrix} -I_p, A \\ B, I_q \end{vmatrix} \begin{vmatrix} -I_p, 0 \\ B, I_q \end{vmatrix} = \begin{vmatrix} I_p+AB, A \\ 0 \quad I_q \end{vmatrix} = |I_p + AB|.$$

Similarly

$$\begin{vmatrix} -I_q, B \\ A, I_p \end{vmatrix} = |I_q + BA|.$$

Q.E.D.

If $q = 1$, then A is a column vector a, B is a row vector b and

$$|I_p + \underline{a}\,\underline{b}'| = 1 + \underline{b}'\underline{a}. \tag{1.1}$$

1.1.1. Minor and Cofactor: For any square matrix $\underline{A} = (a_{ij})$, the minor of the element a_{ij} is the determinant of the matrix formed by deleting the ith row and the jth column of $\underline{A}$. The cofactor A_{ij} of the element a_{ij} of the square matrix $\underline{A}$ is given by $(-1)^{i+j} \times$ the minor of a_{ij}. The determinant of a submatrix of dimension $i \times i$ of a matrix $\underline{A}$ whose diagonal elements are also the diagonal elements of $\underline{A}$ is called a principal minor of order i. The set of leading principal minors is a set of p principal minors of order $1, 2, \ldots, p$ respectively, such that the matrix of the principal minor of order i is a submatrix of the matrix of the principal minor of order $i + 1$.

1.1.2. Inverse Matrix: The inverse of a nonsingular matrix A of dimension $p \times p$ is the unique matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$. Let A_{ij} denote the cofactor of the elements a_{ij} of A. Then

$$A^{-1} = \begin{pmatrix} \frac{A_{11}}{|A|}, & \ldots, & \frac{A_{ip}}{|A|} \\ \cdot, & \ldots, & \cdot \\ \frac{A_{p1}}{|A|}, & \ldots, & \frac{A_{pp}}{|A|} \end{pmatrix}.$$

Theorem 1.2. Let $\underline{A}$ of dimension $p \times p$ and B of dimension $q \times q$ be nonsingular matrices and let U and V be matrices of dimensions $p \times q$ and $q \times p$ respectively. Then

$$(A + UBV)^{-1} = A^{-1} - A^{-1}UB(B + BVA^{-1}UB)^{-1}BVA^{-1}. \tag{1.2}$$

Proof.

$$(A + UBV)[A^{-1} - A^{-1}UB(B + BVA^{-1}UB)^{-1}BVA^{-1}]$$

$$=I - UB(B + BVA^{-1}UB)^{-1}BVA^{-1} + UBVA^{-1} - UBVA^{-1}UB(B + BVA^{-1}UB)^{-1}BVA^{-1}$$

$$=I - (UB + UBVA^{-1}UB)(B + BVA^{-1}UB)^{-1}BVA^{-1} + UBVA^{-1}$$

$$=I - UBVA^{-1} + UBVA^{-1}$$

$$=I.$$

From the uniqueness of the inverse matrix we get the result.

Q.E.D.

If $q = 1$, $B = 1$ and $U = V'$, then

$$(A + UU')^{-1} = A^{-1} - A^{-1}U(1 + U'AU)^{-1}U'A^{-1}$$

$$= A^{-1} - \frac{A^{-1}U\,U'A^{-1}}{1 + U'A^{-1}U}. \qquad (1.3)$$

1.1.3. Rank and Trace: A matrix A can be considered as a set of row vectors or as a set of column vectors arranged in a particular order. The rank of a matrix is defined to be the number of linearly independent rows (which is the same as the number of linearly independent columns) that it contains.

The trace of a square matrix $A = (a_{ij})$ of dimension $p \times p$ is defined by $\operatorname{tr} A = \sum_{i=1}^{p} a_{ii}$.

Obviously for two matrices A,B,

$$\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B,$$

$$\operatorname{tr} AB = \operatorname{tr} BA.$$

For any orthogonal matrix θ,

$$\operatorname{tr} A = \operatorname{tr} A\theta\theta' = \operatorname{tr} \theta'A\theta.$$

1.1.4. Positive Definite Matrix and Quadratic Form: Let $A = (a_{ij})$ be a $p \times p$ matrix and $\underline{x} = (x_1,\ldots,x_p)'$ be a real p-vector. Then $\underline{x}'A\underline{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_ix_j$ is called a quadratic form in the real variables $x_1,\ldots,x_p$.

The matrix A or the associated quadratic form $\underline{x}'A\underline{x}$ is called positive definite if $\underline{x}'A\underline{x} > 0$ for all $\underline{x} \neq \underline{0}$ and is called positive semidefinite (or nonnegative definite) if $\underline{x}'A\underline{x} \geq 0$ for all $\underline{x}$, and $\underline{x}'A\underline{x} = 0$ for at least one $\underline{x} \neq \underline{0}$.

1.1.5. Characteristic Roots and Vectors: These will play a dominant role in the development of multivariate general linear hypotheses.

The characteristic roots of a square matrix A are given by the roots of the characteristic equation

$$|A-\lambda I| = 0 \tag{1.4}$$

where λ is a real constant. Obviously this is an equation of degree p in λ and has exactly p roots. If A is a diagonal matrix, then the diagonal elements are its characteristic roots. Since for any orthogonal matrix θ

$$|A - \lambda I| = |\theta A\theta' - \lambda I|, \tag{1.5}$$

the characteristic roots of the matrix A remain invariant with respect to the transformations

$$A \to \theta A\theta'.$$

The vector $\underline{x} = (x_1, \ldots, x_p)'$, not identically zero, satisfying

$$(A - \lambda I)\,\underline{x} = \underline{0} \tag{1.6}$$

is called a characteristic vector of the matrix A corresponding to the nonzero characteristic root λ. Obviously it follows from (1.6) that if $\underline{x}$ is a characteristic vector of A corresponding to λ, then for any scalar $C(\neq 0)$, $C\underline{x}$ is also a characteristic vector corresponding to λ.

Theorem 1.3. (a) If A is a real symmetric matrix of dimension $p \times p$, then all its characteristic roots are real. (b) The characteristic vectors corresponding to distinct roots of A are orthogonal.

Proof. (a) If possible let λ be a complex characteristic root of A and let $\underline{x} + i\underline{y}$, where $\underline{x} = (x_1, \ldots, x_p)'$, $\underline{y}=(y_1, \ldots, y_p)'$ are real p-vectors, be the complex characteristic vector corresponding to λ. Then from (1.6)

$$(\underline{x} - i\underline{y})'A(\underline{x} + i\underline{y}) = \lambda(\underline{x} - i\underline{y})'(\underline{x} + i\underline{y}) \tag{1.7}$$
$$= \lambda(\underline{x}'\underline{x} + \underline{y}'\underline{y}).$$

Since A is symmetric,

$$(\underline{x} - i\underline{y})'A(\underline{x} + i\underline{y})$$

$$= \underline{x}'A\underline{x} - i\underline{y}'A\underline{x} + i\underline{x}'A\underline{y} + \underline{y}'A\underline{y}$$

$$= \underline{x}'A\underline{x} + \underline{y}'A\underline{y}.$$

As $\underline{x}'A\underline{x} + \underline{y}'A\underline{y}$, $\underline{x}'\underline{x} + \underline{y}'\underline{y}$ are real, we conclude that λ must be real.

(b) Let λ_1, λ_2 be two distinct characteristic roots of A and let $\underline{x}$ and $\underline{y}$ be the characteristic vectors corresponding to λ_1 and λ_2 respectively. Then

$$A\underline{x} = \lambda_1\underline{x}, A\underline{y} = \lambda_2\underline{y}.$$

Hence

$$\underline{x}'A\underline{y} = \lambda_2\underline{x}'\underline{y}, \underline{y}'A\underline{x} = \lambda_1\underline{y}'\underline{x}.$$

Thus $\lambda_1\underline{y}'\underline{x} = \lambda_2\underline{x}'\underline{y}$. Since $\lambda_1 \neq \lambda_2$ we get $\underline{y}'\underline{x} = 0$

Q.E.D.

Let λ be a nonzero characteristic root of a positive definite matrix A and let $x (\neq 0)$ be the corresponding characteristic vector.
Then

$$x'A\underline{x} = \lambda\underline{x}'\underline{x} > 0.$$

This implies that $\lambda > 0$. Thus all characteristic roots of a positive definite matrix are positive.

Example 1.1. Consider the symmetric matrix A of dimension 6×6 given by,

$$\begin{pmatrix} 2.279 & & & & & \\ -.510 & 11.790 & & & & \\ 7.715 & 13.850 & 340.200 & & & \\ .446 & -.235 & 2.604 & .184 & & \\ .310 & .202 & 3.160 & .149 & .500 & \\ 36.730 & -1.340 & 164.400 & 8.775 & 7.333 & 705.1 \end{pmatrix}$$

The ordered characteristic roots of A are

770.290, 277.880, 11.120, 0.469, 0.303, 0.059.

The characteristic vectors of A corresponding to the first three ordered characteristic roots of A are

(.0482, .0048, .3576, .0119, .0104, .9325),
(-.0215, .0504, .9326, -.0026, .0012, -.3568),
(-.0362, .9977, -.0489, -.0218, .0139, .0156).

1.1.6. Factorization of Matrices: We shall state several theorems, mostly without proof, on factorization of matrices which will be useful in the sequel. The reader is referred to Giri (1977) or any standard textbook of matrix theory for proofs.

Theorem 1.4. For every real symmetric matrix A, there exists an orthogonal matrix θ such that $\theta A\theta'$ is a diagonal matrix whose diagonal elements are the characteristic roots of A.

Denoting by $\lambda_1,\ldots,\lambda_p$ the characteristic roots of A, we get

$$|A| = \prod_{i=1}^{p} \lambda_i, \quad \text{tr } A = \sum_{i=1}^{p} \lambda_i.$$

Furthermore if A is positive definite, then $\lambda_i > 0$, $i = 1,\ldots,p$ and hence $|A| > 0$.

Let A be a positive definite matrix of dimension $p \times p$ with characteristic roots $\lambda_1,\ldots,\lambda_p$. Let $D^{\frac{1}{2}}$ be the diagonal matrix with diagonal elements $\lambda_1^{\frac{1}{2}},\ldots,\lambda_p^{\frac{1}{2}}$ and let $D^{\frac{1}{2}}\theta=C$. Then the positive definite matrix A can be factorized as $A = CC'$, where C is a nonsingular matrix and the positive definite quadratic form $\underline{x}'A\underline{x}$ can be transformed to the diagonal form $\underline{y}'\underline{y}$, where $\underline{y} = c\underline{x}$. Furthermore given any positive definite matrix A there exists a nonsingular matrix B $(= C^{-1})$ such that $BAB'=I$. Since

$$x'A^{-1}x = (C^{-1}x)'(C^{-1}x),$$

we conclude that if A is positive definite A^{-1} is also positive definite.

Theorem 1.5. Let A be a symmetric positive definite matrix of dimension $p \times p$ and let B be a $q \times p$ matrix. Then BAB' is symmetric and at least positive semidefinite of the same rank as B.

Theorem 1.6. Let A be symmetric and at least positive semidefinite of dimension $p \times p$ and of rank $r \leq p$. Then A has exactly r positive characteristic roots and the remaining $p-r$ characteristic roots of A are zeros.

Theorem 1.7. Let A be a nonsingular matrix. Then there exists a nonsingular matrix C such that

$$CAC' = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$$

where the order of I equals the number of positive characteristic roots of A and that of $-I$ equals the number of negative characteristic roots of A.

Theorem 1.8. Let A and B be two matrices of dimensions $p \times q$ and $q \times p$ respectively. Then every nonzero characteristic root of AB is also a nonzero characteristic root of BA. The proof follows trivially from the fact that

$$\begin{vmatrix} \lambda I, A \\ B\ , I \end{vmatrix} = \begin{vmatrix} \lambda I, B \\ A\ , I \end{vmatrix} .$$

Theorem 1.9. If A is positive definite and B is positive semidefinite of the same dimension $p \times p$, then there exists a nonsingular matrix C such that $CAC' = I$, $CBC' = D$, where D is a diagonal matrix whose diagonal elements are the roots of $|B - \lambda A| = 0$.

Theorem 1.10. Let A be a matrix of dimension $p \times q$, $p < q$. Then AA' is symmetric and positive semidefinite if the rank of A is less than p and positive definite if the rank of A is equal to p.

Proof. Obviously AA' is symmetric and the rank of AA' is equal to the rank of A. Let the rank of AA' be r. Since AA' is symmetric, there exists an orthogonal $p \times p$ matrix θ such that $\theta AA'\theta'$ is a diagonal matrix with nonzero diagonal elements $\lambda_1, \ldots, \lambda_r$. Writing $\underline{x} = (x_1, \ldots, x_p)'$, $\underline{y} = (y_1, \ldots, y_p)' = \theta\underline{x}$, we get

$$\underline{x}'AA'\underline{x} = \sum_{i=1}^{r} \lambda_i y_i^2 \geq 0$$

for all $\underline{x}$. If $r = p$, then

$$\underline{x}'AA'\underline{x} = \sum_{i=1}^{p} \lambda_i y_i^2 > 0 \text{ for all } \underline{x} \neq \underline{0}.$$

Q.E.D.

Theorem 1.11. Let A and B be two $r \times p$ matrices of the same rank = min (r,p). Then $B'B = A'A$ if and only if there exists an orthogonal matrix θ such that $B = \theta A$.

Proof. If $B = \theta A$, then $B'B = A'\theta'\theta A = A'A$. To prove the converse, consider first the case $r = p$. Since the rank of A is p, the columns of A are linearly independent. Now $B'B = A'A$ implies that $\theta = BA^{-1}$ is an orthogonal matrix and that $B = \theta A$. Consider now the case $r > p$. Since for any two p-dimensional subspaces of the r-space there exists an orthogonal transformation transforming one to the other, we can assume that after a suitable orthogonal transformation, the p column vectors of A, B will lie in the same subspace and the problem is reduced to the case $r = p$.

If $r < p$, we can without any loss of generality assume that the first r column vectors of A are linearly independent. Write $A = (A_1, A_2)$, where A_1 and A_2 are submatrices of dimension $r \times r$ and $r \times (p-r)$ respectively, and similarly $B = (B_1, B_2)$. Since $B'B = A'A$, we get

$$B_1'B_1 = A_1'A_1,\ B_1'B_2 = A_1'A_2,\ B_2'B_2 = A_2'A_2. \qquad (1.8)$$

By the previous argument $B_1'B_1 = A_1'A_1$ implies that there exists an orthogonal matrix $\theta = (B_1')^{-1}A_1'$ such that $B_1 = \theta A_1$. Also $B_1'B_2 = A_1'A_2$ implies that $B_2 = \theta A_2$. Obviously $B_2'B_2 = A_2'A_2$ with $B_2 = \theta A_2$.

Q.E.D.

1.1.7. Idempotent Matrix : A square matrix A is called idempotent if $AA = A$.

Theorem 1.12. Let A be an idempotent matrix of dimension $p \times p$. Then,

(a) if A is nonsingular, $A = I$,

(b) if A is symmetric and if $a_{ii} = 0$, all elements of the ith row and the ith column of A are zeros,

(c) the characteristic roots of A are either unity or zero,

(d) rank $(A) = \text{tr } A$,

(e) if for any matrix B, $A + B = I$, B is idempotent and $AB = BA = 0$.

Proof.

(a) $AA = A$ and A is nonsingular implies that $A = A^{-1}AA = A^{-1}A = I$.

(b) $AA = A$ $(A = (a_{ij}))$ implies that

$$a_{ii} = \sum_{j=1}^{p} a_{ij}a_{ji}.$$

Thus if $a_{ii} = 0$ then $a_{ij} = 0$, $j = 1,\ldots,p$. But A is symmetric, so $a_{ji} = 0$, $j = 1,\ldots,p$.

(c) Let λ be any characteristic root of A and let $\underline{x} \neq \underline{0}$ be the corresponding characteristic vector of A. Then

$$A\underline{x} = \lambda\underline{x},\ AA\underline{x} = \lambda A\underline{x} = \lambda^2\underline{x}.$$

So $\lambda^2\underline{x} = \lambda\underline{x}$. Since $\underline{x} \neq 0$, we conclude that $\lambda^2 = \lambda$, or $\lambda = 1$ or 0.

(d) It follows trivially from (e).

(e) It is trivial.

Q.E.D.

1.1.8. Generalized Inverse (g-inverse): If the matrix A is not a square matrix or if A is a singular square matrix then, obviously, A^{-1} does not exist. However it is possible to define a generalization of the usual inverse A^{-1} by A^{-} in the following way:

Definition 1.1. (g-inverse): Let A be a $p \times q$ matrix of nonzero rank. A g-inverse of A is a $q \times p$ matrix, denoted by A^-, such that for any p-vector $\underline{y}$ and q-vector $\underline{x}$, $A\underline{x} = \underline{y}$ is a consistent equation and $\underline{x} = A^-\underline{y}$ is a solution.

A^- may exist even when A^{-1} does not and it equals A^{-1} whenever A^{-1} exists. It is not necessarily unique, although it is possible, by imposing additional conditions on A^-, to define it uniquely.

Theorem 1.13. Let A be a $p \times q$ matrix. Then A^- exists if and only if $AA^-A = A$.

Proof. Choose $\underline{y}$ as the ith column vector $\underline{a}_i$ of A. Obviously $A\underline{x} = \underline{a}_i$ is consistent and $\underline{x} = A^-\underline{a}_i$ is a solution. Thus

$$AA^-\underline{a}_i = A\underline{x} = \underline{a}_i$$

for all i. This implies that $A = AA^-A$. Conversely if $AA^-A = A$ and $A\underline{x} = \underline{y}$ is consistent, then

$$AA^-A\underline{x} = A\underline{x} = \underline{y}\,, \quad \text{or} \quad AA^-\underline{y} = \underline{y}.$$

Hence $\underline{x} = A^-\underline{y}$ is a solution.

Q.E.D.

Let $H = A^-A$. Then by the above theorem

$$HH = A^-AA^-A = A^-A = H. \tag{1.9}$$

Thus H is an idempotent matrix. Furthermore

$$\text{rank } H = \text{rank } A. \tag{1.10}$$

In a similar fashion we can verify the following theorem.

Theorem 1.14. (a) A general solution of $A\underline{x} = \underline{0}$ is $(H - I)\underline{Z}$ where $\underline{Z}$ is arbitrary. (b) A general solution of the consistent equation $A\underline{x} = \underline{y}$ is $A^-\underline{y} + (H - I)\,\underline{Z}$ where $\underline{Z}$ is arbitrary.

Theorem 1.15. Let A^- be the g-inverse of A. Then rank A = rank A^-.

Proof. Since the rank of A is not zero, there exist nonsingular matrices B, C such that BAC is a diagonal matrix D (not necessarily a square matrix). Denote by D^- the diagonal matrix obtained from D by replacing its nonzero diagonal elements by their reciprocals. Obviously,

$DD^-D = D$, $D^-DD^- = D$. Consider $\underline{A}^- = CD^-B$.

Then

$AA^-A = B^{-1}DC^{-1}\ CD^-BB^{-1}DC^{-1} = B^{-1}DC^{-1} = A$.

By Theorem 1.13, A^- is a generalized inverse of A.

Obviously, rank $(A) \geq$ rank A^-. Now

$$
\begin{aligned}
A^-AA^- &= CD^-BB^{-1}DC^{-1}CD^-B \\
&= CD^-DD^-B \\
&= CD^-B = A^{-1}.
\end{aligned}
$$

Thus $(A^-)^- = A$ and rank $A^- \geq$ rank A. Hence the theorem.

Q.E.D.

1.1.9. Linear Equations: Consider the following system of p linear equations in q unknowns $x_1, \ldots, x_q$ (reals)

$$a_{11}x_1 + \ldots + a_{1q}x_q = b_1$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{p1}x_1 + \ldots + a_{pq}x_q = b_p$$

where a_{ij} and b_i are known real constants. In matrix notation (1.11) can be written as

$$\underline{Ax} = \underline{b}, \tag{1.11}$$

where $A = (a_{ij})$ is a matrix of dimension $p \times q$ and $\underline{x} = (x_1, \ldots, x_q)'$, $\underline{b} = (b_1, \ldots, b_p)'$. If $\underline{b} = \underline{0}$, (1.12) is called a "homogeneous system" of linear equations. Otherwise, it is called a "nonhomogeneous system".

A homogeneous system of linear equations $A\underline{x} = \underline{0}$ always has a solution, $\underline{x} = \underline{0}$ being a trivial solution. This system has a nontrivial solution $\underline{x} \neq \underline{0}$ if and only if the q column vectors of A are linearly dependent. This follows from the fact that $A\underline{x} = \underline{0}$ can be written as

$$x_1\underline{\beta}^1 + \ldots + x_q\underline{\beta}^q = \underline{0} \tag{1.12}$$

where $\underline{\beta}^j$ denotes the jth column vector of A.

Theorem 1.16. The solutions of a homogeneous system of linear equations $A\underline{x} = \underline{0}$ form a vector space.

In other words, if $\underline{\alpha}, \underline{\beta}$ are two solutions of $A\underline{x} = \underline{0}$, then $\underline{\alpha}+\underline{\beta}$ and $C\underline{\alpha}$, where C is a scalar constant, are also solutions of $A\underline{x} = \underline{0}$.

Let X denote the vector space of all solutions of $A\underline{x} = \underline{0}$ and let $R(A)$ denote the vector space spanned by the rows of A.

Theorem 1.17. X is orthogonal to $R(A)$ and rank of X + rank of $A = q$.

Now we consider the nonhomogeneous system of linear equations $A\underline{x} = \underline{b}$ $(\underline{b} \neq \underline{0})$. We first observe that this system may not have any solution. Write $A\underline{x} = \underline{b}$ as

$$x_1\underline{\beta}^1 + \cdots + x_q\underline{\beta}^q = \underline{b} \tag{1.13}$$

where $\underline{\beta}^i$ is the ith column vector of A. From (1.13) it is quite evident that $A\underline{x} = \underline{b}$ has a solution if and only if $\underline{b}$ is linearly dependent on $\underline{\beta}^i$, $i = 1,\ldots,q$. This implies that the rank of A is the same as the rank of (A,b), which is the matrix A augmented by the addition of the column $\underline{b}$.

Let $\underline{\alpha} = (\alpha_1,\ldots,\alpha_q)'$, $\underline{\beta} = (\beta_1,\ldots,\beta_q)'$ be any two solutions of $A\underline{x} = \underline{b}$. Obviously $\underline{\alpha}-\underline{\beta}$ is a solution of $A\underline{x} = \underline{0}$. On the other hand, if α is any particular solution of $A\underline{x} = \underline{b}$ and $\underline{\beta}$ is any solution of $A\underline{x} = \underline{0}$ then $\underline{\alpha} + \underline{\beta}$ is a solution of $A\underline{x} = \underline{b}$. Thus we get the following theorem.

Theorem 1.18. All solutions of $A\underline{x} = \underline{b}$ (if there be any) are obtained by adding to a particular solution of $A\underline{x} = \underline{b}$ all solutions of $A\underline{x} = \underline{0}$.

For finding a particular solution of $A\underline{x} = \underline{b}$ consider the associated homogeneous system

$$A\underline{x} - \underline{b}z = \underline{0}, \tag{1.14}$$

obtained by introducing an additional dummy variable z. It can be easily verified that if the nonhomogeneous system $A\underline{x} = \underline{b}$ admits of a solution then the homogeneous system $A\underline{x} - \underline{b}z = \underline{0}$ admits of a solution $(x_1,\ldots,x_q,z)$ with $z \neq 0$. Such a solution can be found by sweeping out the matrix $B = (A,\underline{b})$. For details the reader is referred to Giri (1974). Write

$y_i = \frac{x_i}{z}$, $i = 1,\ldots,q$.

Clearly, $\underline{y}=(y_1,\ldots,y_q)'$ is a particular solution of $A\underline{x} = \underline{b}$.

If $p = q$ and A is a nonsingular matrix of dimension $p \times p$, then the nonhomogeneous system $A\underline{x} = \underline{b}$ has a unique solution $\underline{x} = A^{-1}\underline{b}$. For the case where A is a singular matrix or the general case where A is not a square matrix we can express a particular solution of $A\underline{x} = \underline{b}$ in terms of a "generalized inverse" of the matrix A.

The concept of the generalized inverse of a matrix A is connected with that of the echelon form (sweep-out form) and the Hermite canonical form of the matrix A.

Echelon form. A matrix A is said to be in an echelon form (or the sweep-out form) if and only if

1) each row of A consists of zeros only or a unity preceded by zeros and not necessarily succeeded by zeros.

2) any column of A containing the first unity of a row has the rest of its elements zeros.

Hermite canonical form. A matrix A is said to be in a Hermite canonical form if its principal diagonal consists of only zeros and unities such that when a diagonal element is zero all the elements in the corresponding row are zeros, and when a diagonal element is unity all the remaining elements in the corresponding column are zeros.

It is now obvious that a matrix in a Hermite canonical form is obtained from a matrix in an echelon form by an interchange of rows.

Theorem 1.19. Let A be a matrix of dimension $p \times q$. Then there exists a nonsingular matrix D of dimension $p \times p$ or a nonsingular matrix E of dimension $q \times q$ such that $DA = H$ or $AE = H$, where H is a matrix in the Hermite canonical form. If, in particular, A is a square matrix of dimension $p \times p$, then H is a square matrix of the same dimension and $HH = H$. In other words H is an idempotent matrix.

Let the rank of A be r. By Theorem 1.19 there exists a nonsingular matrix D of dimension $p \times p$ such that

$$DA = H = \begin{pmatrix} I & C \\ 0 & 0 \end{pmatrix}$$

where I is the unit matrix of dimension $r \times r$ and C is of dimension $r \times (q-r)$.

Let the matrix D be partitioned as

$$D = \begin{pmatrix} D_{(1)} \\ D_{(2)} \end{pmatrix}$$

where $D_{(1)}$ is made up of the first r rows of D and $D_{(2)}$ is made up of the remaining $(p-r)$ rows of D. The nonhomogeneous system $Ax = \underline{b}$ admits of a solution if $D_{(2)}\underline{b} = \underline{0}$ and the Hermite canonical form of the matrix $(A, -\underline{b})$ of the homogeneous system (1.14) is

$$G = \begin{pmatrix} I & C & -D_{(1)}\underline{b} \\ 0 & 0 & \underline{0} \end{pmatrix}$$

The solutions of (1.14) are given by (see Giri, 1974)

$$\underline{x}_{(1)} = -C\underline{x}_{(2)} + (D_{(1)}, \underline{b})z \tag{1.15}$$

where $\underline{x}_{(1)} = (x_1, \ldots, x_r)'$, $\underline{x}_{(2)} = (x_{r+1}, \ldots, x_q)'$. For $\underline{x}_{(2)} = \underline{0}$ and $z = 1$, a particular solution of the nonhomogeneous system $A\underline{x} = \underline{b}$ can be written as

$$\underline{y} = \begin{pmatrix} D_{(1)}\underline{b} \\ \underline{0} \end{pmatrix} = D^*\underline{b}$$

where D^* is a matrix of dimension $q \times p$ such that $D^*_{(1)} = D_{(1)}$ and $D^*_{(2)}\underline{b} = \underline{0}$ and

$$D^* = \begin{pmatrix} D^*_{(1)} \\ D^*_{(2)} \end{pmatrix}$$

Central Library
B11371
519.538 GIR(ANA)
St. Xavier's College

Then D^* is a generalized inverse of A.

Example 1.2. Let

$$A = \begin{pmatrix} 1 & -2 & -3 \\ -2 & 5 & -3 \\ 1 & -3 & 6 \end{pmatrix} \; i$$

Since

last row = -(first row + second row),

A is a singular matrix and hence A^{-1} does not exist. Consider the following system of nonhomogeneous equations in x_1, x_2, x_3

$$\begin{pmatrix} 1 & -2 & -3 \\ -2 & 5 & -3 \\ 1 & -3 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Sweeping out the first and the second columns of $\underline{A}$, we get,

$$\begin{pmatrix} 1 & 0 & -21 \\ 0 & 1 & -9 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5y_1 + 2y_2 \\ 2y_1 + y_2 \\ y_1 + y_2 + y_3 \end{pmatrix}$$

Thus the equations are solvable if and only if $y_1 + y_2 + y_3 = 0$. For any y_1, y_2, y_3 for which the nonhomogeneous equations are solvable we get a particular solution by taking $x_3 = 0$, and this solution is

$$x_1 = 5y_1 + 2y_2, \quad x_2 = 2y_1 + y_2, \quad x_3 = 0,$$

which can be expressed as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} .$$

Thus the matrix

$$\begin{pmatrix} 5 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is a generalized inverse of A. As $y_1+y_2+y_3 = 0$ we could write the particular solution as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

so that the matrix on the right is another generalized inverse of A. Alternatively, we could put $x_3=y_3$ to get a different particular solution as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 21 \\ 2 & 1 & 9 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

so that the matrix on the right gives another generalized inverse of A.

Q.E.D.

1.1.10. Partitioned Matrices: Let a square matrix A be partitioned into

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11}, A_{22} are square submatrices of A and A_{22} is nonsingular.

Theorem 1.20. $|A| = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}|$.

Proof.

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{vmatrix}$$

$$= \begin{vmatrix} A_{11} - A_{12}\,A_{22}^{-1}A_{21} & A_{12} \\ 0 & A_{22} \end{vmatrix}.$$

Central Library
B11371

$$= |A_{22}| \; |A_{11} - A_{12} A_{22}^{-1} A_{21}|$$

Theorem 1.21. A symmetric matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

of dimension $p \times p$ (A_{11} is of dimension $q \times q$) is positive definite if and only if A_{11}, $A_{22} - A_{21} A_{11}^{-1} A_{12}$ are positive definite.

Proof. Let $\underline{x}' = (\underline{x}'_{(1)}, \underline{x}'_{(2)})$, where $\underline{x}'_{(1)} = (x_1, \ldots, x_q)$ is a p-vector

$$\underline{x}'A\underline{x} = (\underline{x}_{(1)} + A_{11}^{-1}A_{12}\underline{x}_{(2)})'A_{11}(\underline{x}_{(1)} + A_{11}^{-1}A_{12}\underline{x}_{(2)})$$

$$+ \underline{x}'_{(2)}(A_{22} - A_{21}A_{11}^{-1}A_{12})\,\underline{x}_{(2)}. \tag{1.16}$$

If A is positive definite, then obviously A_{11}, A_{22} are positive definite. From (1.11) if A_{11}, $A_{22} - A_{21}A_{11}^{-1}A_{12}$ are positive definite, then A is positive definite. Conversely, if A and consequently A_{11} are positive definite, then by taking $\underline{x}(\neq \underline{0})$ such that $\underline{x}_{(1)} + A_{11}^{-1}A_{12}\underline{x}_{(2)} = \underline{0}$, we conclude that $A_{22} - A_{21}A_{11}^{-1}A_{12}$ is positive definite. Q.E.D.

Theorem 1.22. Let a positive definite matrix A be partitioned into submatrices A_{ij}, $i,j = 1,2$ where A_{11} is a square submatrix and let $A^{-1} = B$ be similarly partitioned into submatrices B_{ij}, $i,j = 1,2$. Then

$$A_{11}^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}, \quad A_{22}^{-1} = B_{22} - B_{21}B_{11}^{-1}B_{12},$$

$$A_{11}^{-1}A_{12} = -B_{12}B_{22}^{-1}.$$

The proof is obvious.

1.1.11. Some Results on Derivative: Let $\underline{x} = (x_1, \ldots, x_p)'$ and let

$$\frac{\partial}{\partial \underline{x}} = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_p}\right)'$$

For any scalar function $f(\underline{x})$ of the vector $\underline{x}$

$$\frac{\partial f}{\partial \underline{x}} = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_p} \right)'.$$

Let $A = (a_{ij})$ be a matrix of dimension $p \times p$ and let $f(\underline{x}) = \underline{x}'A\underline{x}$.

Since

$$\underline{x}'A\underline{x} = a_{ii}x_i^2 + 2x_i \sum_{i \neq j} a_{ij}x_j + \sum_{\substack{i \neq k \\ i \neq \ell}} a_{k\ell}x_k x_\ell,$$

we get

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = 2 \sum_{j=1}^{p} a_{ij}x_j.$$

Hence

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = 2A\underline{x}.$$

Similarly one can verify that

$$\frac{\partial |A|}{\partial a_{ii}} = A_{ii}, \quad \frac{\partial |A|}{\partial a_{ij}} = A_{ij},$$

where A_{ij} is the cofactor of the element a_{ij} of A. If A is symmetric then

$$\frac{\partial |A|}{\partial a_{ii}} = A_{ii}, \quad \frac{\partial |A|}{\partial a_{ij}} = 2A_{ij}.$$

1.1.12. Complex Matrices: Here we shall discuss briefly complex matrices - matrices with complex elements and state some results without proof. These results will be useful to study the complex analogue of analysis of variance. The reader is referred to MacDuffee (1948) for proofs and related results. The adjoint A^* of a complex matrix $A = (a_{ij})$ of dimension $p \times q$ is a $q \times p$ matrix $A^* = (\bar{a}_{ij})'$ where " - " denotes the complex conjugate of a_{ij} and " ' " denotes the transpose. For any two complex matrices A, B for which AB is defined, $(AB)^* = B^*A^*$. A complex square matrix A is called unitary if $AA^* = I$, the real identity matrix, and it is called hermitian if $A = A^*$. A square complex matrix A is called normal if $AA^* = A^*A$. A hermitian matrix A of dimension $p \times p$ is called positive definite (semipositive definite) if for all complex p-vectors, (nonnull)

$\underline{\xi}$, $\underline{\xi}^*A\underline{\xi} > 0$ (≥ 0). Since for any hermitian matrix A, $(\underline{\xi}^*A\underline{\xi})^* = \underline{\xi}^*A\underline{\xi}$, the hermitian quadratic fórm assumes only real values.

Theorem 1.23. Let A be a hermitian matrix of dimension $p \times p$. Then

(a) all roots of the characteristic equation $|A - \lambda I| = 0$ are real,

(b) if A is positive definite all characteristic roots of A are positive,

(c) there exists an unitary matrix U such that U^*AU is a diagonal matrix whose diagonal elements are the characteristic roots of A.

Theorem 1.24. Let A be a hermitian positive definite matrix. Then

(a) A can be uniquely factorized as $A = BB^*$ where B is a hermitian positive definite matrix,

(b) there exists a complex nonsingular matrix B such that $BAB^* = I$.

1.1.13. Jacobian of Some Transformations: We merely state some useful results. For proofs the reader is referred to Giri (1976).

Theorem 1.25. Let V be a vector space of dimension p. For $\underline{x}, \underline{y} \in V$, the jacobian of the linear transformation $\underline{y} = (y_1, \ldots, y_p)' = A\underline{x}$ where A is a nonsingular matrix of dimension $p \times p$, is

$$\left|\frac{\partial \underline{y}}{\partial \underline{x}}\right| = \left|\frac{\partial y_1}{\partial \underline{x}}, \ldots, \frac{\partial y_p}{\partial \underline{x}}\right| = |A|.$$

Theorem 1.26. Let the $p \times N$ matrix X be transformed to the $p \times N$ matrix $Y = AX$ where A is a $p \times p$ nonsingular matrix, then the jacobian of the transformation $X \to Y$ is given by

$$\frac{\partial Y}{\partial X} = |A|^N.$$

Theorem 1.27. Let the $p \times N$ matrix X be transformed to the $p \times N$ matrix $Y = A \times B$, where A and B are nonsingular matrices of dimensions $p \times p$ and $N \times N$ respectively. Then the jacobian of the transformation $Y \to X$, is given by

$$\left|\frac{\partial Y}{\partial X}\right| = |A|^N |B|^p.$$

Theorem 1.28. Let A be a symmetric positive definite matrix of dimension $p \times p$. The jacobian of the transformation $CSC' \to S$, where C is a nonsingular matrix of dimension $p \times p$ is $|C|^{p+1}$.

Note that if J is the jacobian of the transformation $Y \to X$, then the jacobian of the inverse transformation $X \to Y$ is J^{-1}.

1.2. GROUPS

The univariate and the multivariate analysis of variance models exhibit symmetries with respect to certain groups of transformations. Solutions to these problems can be obtained in a straightforward way by direct application of group theoretic results. This section will deal with a short review of relevant results of group theory.

A group G is a set of elements associated with an operation τ satisfying the following axioms:

A_1: For any two elements g_1, g_2 in G, $g_1 \tau g_2$ belongs to G.

A_2: For any three elements g_1, g_2, g_3 in G

$$g_1\tau(g_2\tau g_3) = (g_1\tau g_2)\tau g_3.$$

A_3: (Identity element): There exists an element θ in G such that for all g in G, $g\tau\theta$ belongs to G.

A_4: (Inverse element): For every element g in G there exists an element g^{-1} in G such that $g\tau g^{-1} = \theta$.

The following properties of the group G are easy to establish.

For $g \in G$

1. $g\tau g^{-1} = g^{-1}\tau g = \theta$.

2. $g\tau\theta = \theta\tau g = g$.

3. For g, h belonging to G, the equation $g\tau x = h$ in x has a unique solution $x = g^{-1}\tau h$.

For the sake of convenience we will omit τ and write g_1g_2 for $g_1\tau g_2$.

The notation should not be confused with the arithmetical product g_1g_2.

Abelian group. A group G is called abelian if for every pair of elements g_1, g_2 belonging to G, $g_1\tau g_2$ belongs to G.

Example 1.3. The set of all real numbers R under addition forms an abelian group with zero as the identity element and with $g^{-1} = -g$.

Permutation group. Let χ be a set and let G be the set of all one to one mappings $g : \chi$ onto χ such that (i) if $gx = gy$ for x, y belonging to χ then $x = y$, and (ii) for $x, y \in \chi$ there exists a $g \in G$ satisfying $y = gx$. With the group operation defined by, $g_1g_2x = g_1(g_2x)$ for $g_1, g_2 \in G$, $x \in \chi$, G forms a group, and is called the permutation group.

Translation group. Let χ be a linear space of dimension n (that is, if x and y are in χ so is $ax + by$ for any scalars a, b). For every $x_1 \in \chi$, define $g_{x_1}(x) = x + x_1$, $x \in \chi$. The set of all g_{x_1} forms an additive abelian group and is called the translation group.

Full linear group. Let χ be a linear subspace of dimension n and let $G_\ell(n)$ be the set of all nonsingular linear transformations of χ onto χ. Then $G_\ell(n)$ forms a group. The group operation is the matrix multiplication, the identity element being the identity matrix.

Affine group. Let χ be a linear space of dimension n. An element of the affine group $G = (G_\ell(n), \chi)$ is the pair (g, x), $g \in G_\ell(n)$, $x \in \chi$. The group operation for the affine group G is given by,

$$(g_1, x_1)(g_2, x_2) = (g_1g_2, gx_2 + x_1)$$

for $g_1, g_2 \in G_\ell(n)$; $x_1, x_2 \in \chi$. The identity element of the affine group G is $(I, 0)$ and the inverse element is given by

$(g,x)^{-1} = (g^{-1}, -g^{-1}x)$.

1.2.1. Subgroup: Let G be a group and the group operation in G be τ. If the restriction of τ to a nonempty subset H of G also satisfies all the group axioms A_1-A_4, then H is called a subgroup of G.

Theorem 1.29. Let G be a group and $H \subset G$. Then H is a subgroup of G if and only if (i) H is nonempty (ii) if $g_1, g_2 \in H$ then $g_1 g_2^{-1} \in H$.

Proof. By (i) and (ii) if g_1, g_2 are in H so are $g_1 g_1^{-1} = \theta$ and $g_2^{-1} = \theta g_2^{-1}$. Thus $g_1(g_2^{-1})^{-1} = g_1 g_2$ is also in H. The axiom A_3 is true in H as it is true in G. Hence H is a group. Conversely if H is a group, then clearly H satisfies (i) and (ii) above.

Q.E.D.

1.2.2. Normal Subgroup: A subgroup H of G is a normal subgroup if for all $h \in H$ and $g \in G$

$ghg^{-1} \in H$,

or equivalently $gHg^{-1} = H$.

Example 1.4. Let χ be a linear space of dimension n and let $\underline{I}$ be the identity matrix of dimension $n \times n$. For $g \in G_\ell(n)$, $x \in \chi$,

$$(g,x)(I,x)(g,x)^{-1} = (g,x)(I,x)(g^{-1},-g^{-1}x)$$

$$= (g,x)(g^{-1},-g^{-1}x + x)$$

$$= (I,gx).$$

Since (I,gx) is an element of (I,χ), we conclude that (I,χ) is a normal subgroup of $(G_\ell(n),\chi)$.

1.2.3. Quotient Group: Let G be a group and let H be a normal subgroup of G. Define $G|\hat{H}$ as the set of all elements $gH = \{gh | h \in H\}$, $g \in G$. Noting that for $g_1, g_2 \in G$, $(g_1H)(g_2H)$ is the set of all elements obtained by multiplying, with respect to the group operation, all elements of g_1H with all elements of g_2H, we see that $G|H$ forms a group, it is called the quotient group of $G(\mathrm{mod}\ H)$.

The fact that $G|H$ satisfies the axioms A_1-A_4 can be verified from the following:

(i) $g_1H = g_2H$

<=> $g_2^{-1}g_1H = H$

<=> $g_2^{-1}g_1 \in H$

<=> $g_1 \in g_2H$.

(ii) Since H is a normal subgroup, $g_2Hg_2^{-1} = H$ or $g_2H = Hg_2$ and $(g_1H)(g_2H) = g_1Hg_2H = g_1g_2H$.

Hence g_1g_2H is in $G|H$.

(iii) H is the identity element in $G|H$.

(iv) Since

$gHg^{-1}H = gg^{-1}HH = gg^{-1}H = H$

$(gH)^{-1} = g^{-1}H$.

1.2.4. Homomorphism: Let G, H be two groups. Then a mapping f of G into H is called a homomorphism if it preserves the group operations, that is, for g_1 and g_2 in G,

$f(g_1g_2) = f(g_1)f(g_2)$.

The product g_1g_2 is defined with respect to the group operation in G and $f(g_1)f(g_2)$ is defined with respect to the group operation in H.

Theorem 1.30. (a) If θ is the identity element of G, $f(\theta)$ is the identity element of H.

(b) For $g \in G$, $f(g^{-1}) = [f(g)]^{-1}$.

Proof.

(a) $f(g) = f(g\theta) = f(g)f(\theta)$

(b) $f(\theta) = f(gg^{-1}) = f(g)f(g^{-1})$.

Q.E.D.

1.2.5. Direct Product: Let G, H be groups and let $G \times H$ be the cartesian product of G and H. With the operation (for

g_1, g_2, g_1g_2 are in G and h_1, h_2, h_1h_2 are in H),

$$(g_1, h_1)(g_2, h_2) = (g_1g_2, h_1h_2).$$

G × H forms a group and is called the direct product of G and H.

1.2.6. Transformation Group: Let χ be a set and let G be a group. G operates from the left on χ if there exists a function on $G \times \chi$ into χ whose value at (g, x), $g \in G$, $x \in \chi$ is given by gx and if it satisfies

(i) $\theta x = x$ for all $x \in \chi$, where θ is the identity element of G.

(ii) $g_1g_2x = g_1(g_2x)$.

If for every pair $(x_1, x_2) \in \chi$ there exists a $g \in G$ such that $gx_1 = x_2$ then we say G operates transitively on χ from the left. It follows from (i) and (ii) that if for $g \in G$ and $x_1, x_2 \in \chi$,

$$gx_1 = gx_2 = y,$$

then

$$g^{-1}gx_1 = g^{-1}gx_2 = g^{-1}y,$$

and $x_1 = x_2 = g^{-1}y$.

Example 1.5. Let χ be a linear space of dimension n. Then $G_\ell(n)$ operates transitively on $\chi - \{0\}$.

EXERCISE 1

1. Show that for any two matrices A, B

   $$(A + B)(A - B) = AA - BB$$

   if and only if AB=BA.

2. Show that for any three matrices A,B,C, rank (AB) + rank(BC) ≤ rank(A) + rank(ABC).

3. Show that for any matrix A, AA is symmetric if A is skew.

4. A skew matrix with an odd number of rows is singular.

5. The determinant of a skew-symmetric matrix is nonnegative.

6. For any nonsingular matrix A, show that all its principal minors are nonsingular.

7. Let A be a nonsingular matrix. Show that

(i) if A is symmetric and positive definite, then there exists a unique lower triangular matrix T with positive diagonal elements such that $A = TT'$.

(ii) There exists a lower triangular nonsingular matrix C such that $A = CC'$.

8. If A is a square matrix, then there exists an orthogonal matrix 0 such that $0A$ is a lower triangular matrix.

9. Let A, B be two positive definite matrices of the same dimension.

Show that

$$|\theta A + (1-\theta)B| \geq |A|^{\theta}|B|^{1-\theta}, \quad 0 \leq \theta \leq 1.$$

10. Let $A_1,\ldots,A_k$ be symmetric matrices of the same dimension. If $A_iA_j = 0$ for $i \neq j$ and if $\sum_{i=1}^{k} A_i$ is an idempotent matrix, then show that A_i for each i is an idempotent matrix and $\text{rank}\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \text{rank}\,(A_i)$.

11. Let A be a symmetric positive definite matrix and let B be a symmetric semipositive definite matrix of the same dimension $p \times p$ and rank $r \leq p$. Show that

(i) all roots of the characteristic equation $|B - \lambda A| = 0$ are zeros if and only if $B = 0$,

(ii) all roots of $|B - \lambda A| = 0$ are unity if and only if $B = A$.

12. Let

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{pmatrix} \sigma^2, \quad 0 < \rho \leq 1.$$

Show that the largest characteristic root of Σ is $\lambda_1 = \sigma^2(1 + (p-1)\rho)$, and the corresponding normalized characteristic vector of Σ is $(1/\sqrt{p}, \ldots, 1/\sqrt{p})$.

13. Show that for any lower triangular matrix A the diagonal elements are its characteristic roots.

14. Show that an orthogonal transformation can be treated as a change of axes about a fixed origin.

15. Let A, B be two matrices of the same dimension $p \times q$. Then show that

(a) $\frac{\partial(\text{tr } A'B)}{\partial B} = A = \frac{\partial(\text{tr } AB')}{\partial B}$

where

$\frac{\partial}{\partial B} = \left(\frac{\partial}{\partial b_{ij}}\right)$, $B = (b_{ij})$,

(b) if A is a square symmetric matrix, then

$\frac{\partial(\text{tr } AA)}{\partial A} = 2A.$

16. Show that the set of all orthogonal matrices of the same dimension forms a group.

17. Show that the set of all $p \times p$ nonsingular triangular matrices forms a group.

18. Show that the intersection of two subgroups H and K of a group G is also a subgroup of G.

19. Show that every subgroup of an abelian group is a normal subgroup.

REFERENCES

Birkoff, G. and McLane, S. (1953). A survey of modern algebra, Macmillan, N.Y.

Giri, N. (1974). Introduction to probability and statistics, part 2, Statistics, Marcel Dekker, N.Y.

Giri, N. (1977). Multivariate statistical inference, Academic Press, N.Y.

Markus, M. and Mine, H. (1967). Introduction to linear algebra, Macmillan, N.Y.

MacDuffee, C. (1946). The theory of matrices, Chelsea, New York.

Perlis, S. (1952). Theory of matrices, Addison Wesley, Cambridge, Mass.

Rao, C.R. (1965). Linear statistical inference and its applications, Wiley, N.Y.

Roy, S.N. (1957). Some aspects of multivariate analysis, Wiley, N.Y.

CHAPTER 2. BASIC STATISTICAL CONCEPTS

Statistical inference in linear models using analysis of variance techniques depends on the partition of a chisquare variable into several independent components each distributed as a chisquare variable and on F and student-t statistics. In this chapter we shall first discuss the nonnull distributions of these statistics. The multivariate analogues of these distributions will also be considered to facilitate discussion on multivariate analysis of variance. Most analysis of variance test criteria are invariant procedures withrespect to certain groups of transformations leaving the hypothesis in question invariant. Thus to obtain the appropriate test criteria and to study their optimum properties we need to outline in some detail the principle of invariance in statistical inference. This will also be done in this chapter.

2.1. NONCENTRAL CHISQUARE DISTRIBUTION

Let $X_1,\dots,X_n$ be independently distributed normal random variables with $E(X_i) = \mu_i$ $\mathrm{Var}(X_i) = \sigma_i^2$ and let $U = \sum_{i=1}^{N} X_i^2/\sigma_i^2$.

__Theorem 2.1.__ The random variable U has a noncentral chisquare distribution $\chi_N^2(\lambda^2)$ with N degrees of freedom and the noncentrality parameter $\lambda^2 = \sum_{i=1}^{N} \mu_i^2/\sigma_i^2$. Its probability density function is given by

$$f_U(u) = \begin{cases} \dfrac{\exp\{-\frac{1}{2}(\lambda^2+u)\}}{\sqrt{\pi}\,2^{\frac{1}{2}N}} \, u^{\frac{1}{2}N-1} \displaystyle\sum_{j=o}^{\infty} \frac{(\lambda^2 u)^j \, \Gamma(j+\frac{1}{2})}{(2j)!\,\Gamma(\frac{1}{2}N+j)}, & \text{if } u \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

__Proof.__ Let $Y_i = X_i/\sigma_i$, $i = 1,\dots,N$. Obviously Y_i's are independently distributed normal random variables with

$$E(Y_i) = \frac{\mu_i}{\sigma_i} = \nu_i \text{ (say)}$$

and unit variance. Let $\underline{Y} = (Y_1,\dots,Y_N)'$ and let

$\underline{Z} = (Z_1, \ldots, Z_N)' = O\underline{Y}$

where O is an orthogonal matrix of the following form

$$O = \begin{pmatrix} \frac{v_1}{\lambda}, & \ldots, & \frac{v_N}{\lambda} \\ O_{21}, \ldots, & & O_{2N} \\ \ldots, & \ldots, & \ldots \\ O_{N1}, \ldots, & & O_{NN} \end{pmatrix}$$

Then

$$E(Z_i) = \begin{cases} \lambda & \text{if } i = 1, \\ 0 & \text{if } i > 1, \end{cases}$$

$$\text{Cov}(Z_i, Z_{i'}) = \sum_{j=1}^{N} O_{ij}\, O_{i'j}\, \text{Var}(Y_j) + \sum_{j=1}^{N} \sum_{j'=1}^{N} O_{ij}\, O_{i'j'} \text{Cov}(Y_j, Y_{j'})$$

$$= \sum_{j=1}^{N} O_{ij}\, O_{i'j}$$

$$= \begin{cases} 1 & \text{if } i = i' \\ 0 & \text{otherwise} \end{cases}$$

Thus $Z_1, \ldots, Z_n$ are independently distributed normal random variables with unit variance and with $E(Z_1) = \lambda$, $E(Z_i) = 0$, $i > 1$.

Let $X = Z_1^2$, $Y = \sum_{i=2}^{N} Z_i^2$. Then X, Y are independent. The probability density functions $f_X(x)$, $f_Y(y)$ of X and Y respectively are given by,

$$f_X(x) = \frac{1}{2\sqrt{2\pi}} x^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\sqrt{x}-\lambda)^2\} + \frac{1}{2\sqrt{2\pi}} x^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\sqrt{x}+\lambda)^2\}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(x+\lambda^2)\}\, x^{\frac{1}{2}-1} \sum_{j=0}^{\infty} \frac{(\lambda^2 x)^j}{(2j)!}, \qquad (2.2)$$

$$f_Y(y) = \frac{1}{\Gamma(\frac{1}{2}(N-1))\, 2^{\frac{1}{2}(N-1)}} \exp\{-\tfrac{1}{2}y\}\, y^{\frac{1}{2}(N-1)-1},$$

whenever $x \geq 0$, $y \geq 0$. From (2.2), the probability density function of $U = X+Y$ is

$$f_U(u) = \frac{\exp\{-\frac{1}{2}(\lambda^2+u)\}}{\sqrt{\pi}\,2^{\frac{1}{2}N}} u^{\frac{1}{2}N-1} \sum_{j=o}^{\infty} \frac{(\lambda^2)^j u^j \Gamma(j+\frac{1}{2})}{(2j)!\Gamma(\frac{1}{2}N+j)}$$

$$\int_0^1 \frac{\Gamma(\frac{1}{2}N+j)}{\Gamma(j+\frac{1}{2})\Gamma(\frac{1}{2}(N-1))} v^{\frac{1}{2}+j-1}(1-v)^{\frac{1}{2}(N-1)-1}\, dv$$

$$= \frac{\exp\{-\frac{1}{2}(\lambda^2+u)\}}{\sqrt{\pi}\,2^{\frac{1}{2}N}} u^{\frac{1}{2}N-1} \sum_{j=0}^{\infty} \frac{(\lambda^2)^j u^j \Gamma(j+\frac{1}{2})}{(2j)!\Gamma(\frac{1}{2}N+j)}$$

Q.E.D.

where $u \geq 0$, $v = x/u$.

Remarks

1. Since for any positive integer k

$\Gamma(2k+1) = 2^{2k}\,\Gamma(k+1)\,\Gamma(K+\frac{1}{2})$,

we can rewrite $f_U(u)$ as

$$f_U(u) = \sum_{k=o}^{\infty} \frac{\exp\{-\frac{1}{2}\lambda^2\}(\frac{1}{2}\lambda^2)^k}{k!} \frac{\exp\{-\frac{1}{2}u\}\, u^{\frac{1}{2}N+k-1}}{\Gamma(\frac{1}{2}N+k)\;2^{\frac{1}{2}N+k}} . \tag{2.3}$$

$$= \sum_{k=o}^{\infty} p_K(k)\, f_{\chi^2_{N+2k}}(u),$$

where $p_K(k)$ is the probability mass function of the Poisson random variable K with parameter $\frac{1}{2}\lambda^2$ and $f_{\chi^2_{N+2k}}$ is the probability density function of the central chisquare random variable with $N+2k$ degrees of freedom.

2. The characteristic function of U is given by

$$\phi_U(t) = E(e^{itU})$$

$$=(1-2it)^{-\frac{1}{2}N} \exp\{\frac{it\lambda^2}{1-2it}\}, \tag{2.4}$$

for real t and $i = \sqrt{-1}$. From the unicity property of the characteristic function it follows that if $U_1,\ldots,U_k$ are independently distributed noncentral chisquare random variables $\chi^2_{N_i}(\lambda_i^2)$, $i = 1,\ldots,k$, then $\sum_{i=1}^{k} U_i$ is distributed as a noncentral chisquare with $\sum_{i=1}^{k} N_i$ degrees of freedom and with noncentrality parameter $\sum_{i=1}^{k} \lambda_i^2$.

3. From (2.4) it follows that

$$E(\chi_N^2(\lambda^2)) = N+\lambda^2, \quad Var(\chi_N^2(\lambda^2)) = 2N+4\lambda^2. \tag{2.5}$$

2.2. NONCENTRAL F, NONCENTRAL BETA AND NONCENTRAL t-DISTRIBUTIONS

Let the random variable X, distributed as $\chi_m^2(\lambda^2)$ and the random variable Y, distributed as χ_n^2 be independent and let $F=nX/mY$.

Theorem 2.2. The probability density function of F is given by

$$f_F(u)=\begin{cases} \frac{m}{n}\exp\{-\tfrac{1}{2}\lambda^2\} \sum_{j=o}^{\infty} \dfrac{(\frac{1}{2}\lambda^2)^j\,\Gamma(\frac{1}{2}(m+n)+j)\,\{(m/n)u\}^{\frac{1}{2}m+j-1}}{\Gamma(j+1)\Gamma(\frac{1}{2}m+j)\Gamma(\frac{1}{2}n)(1+\frac{m}{n}u)^{\frac{1}{2}(m+n)+j}} & \text{if } u \geq 0. \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Proof. The joint probability density function of X,Y is given by

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\,2^{\frac{1}{2}(m+1)}} \exp\{-\tfrac{1}{2}(x+y+\lambda^2)\}$$

$$x^{\frac{1}{2}m-1}\,y^{\frac{1}{2}n-1} \sum_{j=o}^{\infty} \frac{(\lambda^2 x)^j\,\Gamma(j+\frac{1}{2})}{(2j)!\,\Gamma(j+\frac{1}{2}m)} \tag{2.7}$$

From (2.7) the probability density function of F is

$$f_F(u) = \frac{m}{n\sqrt{\pi}\;2^{\frac{1}{2}(m+n)}\,\Gamma(\frac{1}{2}n)} \exp\{-\tfrac{1}{2}\lambda^2\}$$

$$\sum_{j=o}^{\infty} \frac{(\lambda^2)^j\,\Gamma(j+\frac{1}{2})}{(2j)!\;\Gamma(j+\frac{m}{2})} \left(\frac{m}{n}u\right)^{j+\frac{1}{2}m-1} \int_0^{\infty} y^{\frac{1}{2}(m+n)+j-1} \exp\{-\tfrac{1}{2}y(1+\tfrac{m}{n}u)\}dy$$

$$= \frac{m}{n}\exp\{-\tfrac{1}{2}\lambda^2\} \sum_{j=o}^{\infty} \frac{(\frac{1}{2}\lambda^2)^j\,\Gamma(\frac{1}{2}(m+n)+j)\,(\frac{m}{n}u)^{\frac{1}{2}m+j-1}}{\Gamma(j+1)\Gamma(\frac{1}{2}m+j)\Gamma(\frac{1}{2}n)(1+\frac{m}{n}u)^{\frac{1}{2}(m+n)+j}} \tag{2.7a}$$

provided that $u \geq 0$.

Corollary. Let $W = X/(X+Y)$. The probability density function of W is given by,

$$f_W(\omega) = \sum_{k=o}^{\infty} p_K(k)\, f_{\frac{1}{2}m+k,\frac{1}{2}n}(\omega) \tag{2.8}$$

where $p_K(k)$ is as defined in (2.3) and $f_{\frac{1}{2}m+k,\frac{1}{2}n}(\omega)$ is the probability density function of the central beta random variable with parameter $(\frac{1}{2}m+k, \frac{1}{2}n)$, that is,

$$f_W(\omega) = \frac{\Gamma(\frac{1}{2}(m+n)+k)}{\Gamma(\frac{1}{2}m+k)\,\Gamma(\frac{1}{2}n)}\ \omega^{\frac{1}{2}m+k-1}\ (1-\omega)^{\frac{1}{2}n-1}$$

provided $0 \leq \omega \leq 1$.

The proof follows immediately from Theorem 2.2. The distribution (2.8) is called the noncentral beta distribution.

Theorem 2.3. Let the random variable X, distributed normally with mean μ and variance σ^2, and the random variable Y/σ^2, distributed as a central chisquare with n degrees of freedom, be independent and let $t = \sqrt{n}\, X/\sqrt{Y}$. The probability density function of t is given by

$$f_t(t) = \frac{n^{\frac{1}{2}n}\ \exp\{-\frac{1}{2}\lambda^2\}(n+t^2)^{-\frac{1}{2}(n+1)}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)}$$

$$\sum_{j=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+j+1))\,\lambda^j}{j!}\ \frac{2^{\frac{1}{2}j}\,t^j}{(n+t^2)^{\frac{1}{2}j}}\,, \quad -\infty < t < \infty\,, \qquad (2.9)$$

0 otherwise,

where $\lambda^2 = \mu^2/\sigma^2$.

Proof. Let $X_1 = \sqrt{n}X/\sigma$; $X_2 = Y/\sigma^2$. The joint probability density function of X_1, X_2 is given by

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\sqrt{2n\pi}\,2^{\frac{1}{2}n}\,\Gamma(\frac{1}{2}n)}\ \exp\{-\tfrac{1}{2}n(x_1-\sqrt{n}\lambda)^2\}\exp\{-\tfrac{1}{2}x_2\}x_2^{\frac{1}{2}n-1}.$$

Thus the probability density function of $t=x_1/\sqrt{x_2}$ is given by

$$f_t(t) = \int_o^{\infty} \frac{1}{\sqrt{2n\pi}\ 2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\ \exp\{-\tfrac{1}{2}n(t\sqrt{x_2}-\sqrt{n}\lambda)^2\}e^{-\frac{1}{2}x_2}x_2^{\frac{1}{2}(n-1)}\ dx_2$$

$$= \frac{1}{\sqrt{2n\pi}\ 2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\ \exp\{-\tfrac{1}{2}\lambda^2\}\ \sum_{j=o}^{\infty}\frac{(\frac{t\lambda}{\sqrt{n}})^j}{j!}\ \int_o^{\infty} \exp\{-\tfrac{1}{2}x_2(1+\tfrac{t^2}{n})\}x_2^{\frac{1}{2}(n+1+j)-1}dx_2$$

$$= \frac{n^{\frac{1}{2}n}\exp\{-\frac{1}{2}\lambda^2\}}{\sqrt{\pi}\Gamma(\frac{1}{2}n)}(n+t^2)^{-\frac{1}{2}(n+1)}\ \sum_{j=o}^{\infty}\ \frac{\lambda^j}{j!}\Gamma(\tfrac{1}{2}(n+j+1))\frac{2^{\frac{1}{2}j}\ t^j}{(n+t^2)^{\frac{1}{2}j}}.$$

Q.E.D.

The distribution (2.9) is called the noncentral t-distribution.

2.3. MULTIVARIATE NORMAL DISTRIBUTION, QUADRATIC FORM AND COCHRAN'S THEOREM

The analysis of variance technique involves the decomposition of a quadratic form into several orthogonal quadratic forms.

In this section we shall discuss some important results concerning the distribution of quadratic forms and in particular Cochran's theorem for correlated random variables. Results for uncorrelated random variables which are useful for univariate analysis of variance will follow trivially from these.

We will first outline the basic results concerning the multivariate normal distribution. The reader is referred to Giri (1976) for proofs of these results.

The random vector $\underline{X} = (X_1,\ldots,X_N)'$ with values in E^N (Euclidean N-space) is said to have an N-variate normal distribution with mean vector $E(\underline{X}) = \underline{\mu} = (\mu_1,\ldots,\mu_N)'$ and positive definite covariance matrix $\Sigma = E(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})'$ if its probability density function is given by

$$f_{\underline{X}}(\underline{x}) = (2\pi)^{-\frac{1}{2}N}|\Sigma|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'\}. \tag{2.10}$$

The characteristic function of $\underline{X}$ is

$$E(\exp\{i\underline{t}'\underline{X}\}) = \exp\{i\underline{t}'\underline{\mu} - \tfrac{1}{2}\underline{t}'\Sigma\underline{t}\} \tag{2.11}$$

where $\underline{t} \in E^N$. Since the characteristic function determines uniquely the distribution function, the multivariate normal distribution is completely determined by the mean vector $\underline{\mu}$ and the covariance matrix Σ.

If $\underline{X}$ has an N-variate normal distribution with mean vector $\underline{\mu}$ and positive definite covariance matrix Σ, then every linear combination $\underline{L}'\underline{X}$, $\underline{L} = (\ell_1,\ldots,\ell_N)' \in E^N$, has a univariate normal distribution with mean $\underline{L}'\underline{\mu}$ and variance $\underline{L}'\Sigma\underline{L}$. The converse of this result, that is, if $\underline{X}$ with mean $\underline{\mu}$ and covariance matrix Σ be such that every linear combination of the components of $\underline{X}$ has a univariate normal distribution then $\underline{X}$ has an N-variate normal distribution, also holds. This fact leads to the following general definition of the multivariate normal distribution which covers cases where the covariance matrix is a singular matrix and as a consequence $f_{\underline{X}}(\underline{x})$ as given in (2.10) does not exist.

Definition 2.1. An N-variate random vector $\underline{X}$ with values in E^N is said to have a normal distribution with mean $\underline{\mu}$ and covariance matrix Σ if and only if every linear combination of the components of $\underline{X}$ has a univariate normal distribution.

According to this definition, if $\underline{X}$ has an N-variate normal distribution then each component X_i of $\underline{X}$ is distributed as a univariate normal variable, so that $-\infty < E(X_i)=\mu_i < \infty$ and $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$ for all i and hence $\mathrm{Cov}(X_i, X_j) = \sigma_{ij} < \infty$ for all i, j. Thus $E(\underline{X})$ and $\mathrm{Cov}(\underline{X})$ exist and we denote them by $\underline{\mu}$, Σ respectively. It is not necessary in this definition that Σ is a nonsingular matrix.

From (2.10) it is easy to conclude that if Σ is a diagonal matrix then the components of $\underline{X}$ are independently distributed normal random variables. Let

$$\underline{X} = (\underline{X}_{(1)}, \underline{X}_{(2)})', \quad \underline{X}_{(1)}=(X_1,\ldots,X_q)', \quad \underline{X}_{(2)}=(X_{q+1},\ldots,X_N)', \tag{2.12}$$

and $\underline{\mu}$ be similarly partitioned as $\underline{\mu} = (\underline{\mu}_{(1)}, \underline{\mu}_{(2)})'$ and let Σ be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & , & \Sigma_{(12)} \\ \\ \Sigma_{(21)} & , & \Sigma_{(22)} \end{pmatrix}$$

where $\Sigma_{(ij)} = E(\underline{X}_{(i)} - \underline{\mu}_{(i)})(\underline{X}_{(j)} - \underline{\mu}_{(j)})'$.

Theorem 2.4. Suppose that $\underline{X}$ has an N-variate normal distribution with mean $\underline{\mu}$ and covariance matrix Σ. Then $\Sigma_{(12)} = 0$ if and only if $\underline{X}_{(1)}, \underline{X}_{(2)}$ are independently normally distributed with means $\underline{\mu}_{(1)}, \underline{\mu}_{(2)}$ and covariance matrices $\Sigma_{(11)}, \Sigma_{(22)}$ respectively.

Proof. The characteristic function of any linear combination $\underline{L}'\underline{X}$, $\underline{L} = (\ell_1,\ldots,\ell_N)' \in E^N$, is given by

$$\phi(t,\underline{L}) = \exp\{it\underline{L}'\underline{X}\}$$

$$= \exp\{it\underline{L}'\underline{\mu} - \tfrac{1}{2}t^2\underline{L}'\,\Sigma\,\underline{L}\} .$$

Let $\underline{L} = (\underline{L}_{(1)}, \underline{L}_{(2)})'$, where $\underline{L}_{(1)} = (\ell_1, \dots, \ell_q)'$, and let $\Sigma_{(12)} = 0$. Then

$$\phi(t,\underline{L}) = \prod_{j=1}^{2} \exp\{it\underline{L}'_{(j)}\underline{\mu}_{(j)} - \tfrac{1}{2}t^2\underline{L}'_{(j)}\,\Sigma_{(jj)}\,\underline{L}_{(j)}\}.$$

Since $\underline{L}$ is arbitrary, taking $t = 1$, we conclude that the characteristic function of $\underline{X}$ is the product of the characteristic functions of $\underline{X}_{(1)}$ and $\underline{X}_{(2)}$ and each one is the characteristic function of a multivariate normal distribution, and hence the theorem.

Q.E.D.

Theorem 2.5. Let $\underline{X} = (X_1, \dots, X_N)'$ have an N-variate normal distribution with mean $\underline{\mu}$ and covariance matrix Σ and let $\underline{Y} = C\underline{X}$, where C is a matrix of dimension $q \times N$ of rank q. Then $\underline{Y}$ has a q-variate normal distribution with mean $C\underline{\mu}$ and covariance matrix $C\,\Sigma\,C'$.

Proof. For any nonnull q-vector $\underline{L} \in E^q$, $\underline{L}'\underline{Y} = (\underline{L}'C)\underline{X}$. By definition 2.1, $\underline{L}'C\underline{X}$ and equivalently $\underline{L}'\underline{Y}$ has a univariate normal distribution with mean $\underline{L}'C\,\underline{\mu}$ and variance $\underline{L}'C\,\Sigma\,C'\underline{L}$. Since $\underline{L}$ is arbitrary, $\underline{Y}$ has a q-variate normal distribution with mean $C\underline{\mu}$ and covariance matrix $C\,\Sigma\,C'$.

Q.E.D.

Theorem 2.6. Let $\underline{X}$ have an N-variate normal distribution with mean $\underline{\mu}$ and covariance matrix Σ and let $\underline{X}, \underline{\mu}, \underline{\Sigma}$ be partitioned as in (2.12). Then

(i) the marginal distribution of $\underline{X}_{(1)}$ is q-variate normal with mean $\underline{\mu}_{(1)}$ and covariance matrix $\Sigma_{(11)}$,

(ii) the conditional distribution of $\underline{X}_{(2)}$ given that $\underline{X}_{(1)} = \underline{x}_{(1)}$ is (N-q)- variate normal with mean $\underline{\mu}_{(2)} + \Sigma_{(2)}\,\Sigma^{-}_{(11)}(\underline{x}_{(1)} - \underline{\mu}_{(1)})$ and covariance matrix $\Sigma_{(22)} - \Sigma_{(21)}\,\Sigma^{-}_{(11)}\,\Sigma_{(12)}$.

Proof.

(i) Since every linear combination of the components of $\underline{X}_{(1)}$ is also univariate normal, by definition 2.1, $\underline{X}_{(1)}$ has the q-variate normal distribution with mean $\underline{\mu}_{(1)}$ and covariance matrix $\Sigma_{(11)}$.

(ii) Let us first observe that any vector orthogonal to the columns of $\Sigma_{(11)}$ is also orthogonal to the columns of $\Sigma_{(12)}$.

So the linear space spanned by the columns of $\Sigma_{(12)}$ is a subspace of the linear space spanned by the columns of $\Sigma_{(11)}$. Hence there exists a matrix A such that

$$\Sigma_{(21)} = A\,\Sigma_{(11)}. \tag{2.13}$$

Now by using the property $\Sigma_{(11)}\,\Sigma^{-}_{(11)}\,\Sigma_{(11)} = \Sigma_{(11)}$ for the generalized inverse $\Sigma^{-}_{(11)}$ of $\Sigma_{(11)}$ we obtain

$$\Sigma_{(21)}\Sigma^{-}_{(11)}\,\Sigma_{(11)} = A\,\Sigma_{(11)}\,\Sigma^{-}_{(11)}\,\Sigma_{(11)} = A\,\Sigma_{(11)} = \Sigma_{(21)}. \tag{2.14}$$

Thus the covariance of $\underline{X}_{(2)} - \underline{\mu}_{(2)} - \Sigma_{(21)}\,\Sigma^{-}_{(11)}\,(\underline{X}_{(1)} - \underline{\mu}_{(1)})$ is equal

$$E(\underline{X}_{(2)} - \underline{\mu}_{(2)} - \Sigma_{(21)}\Sigma^{-}_{(11)}\,(\underline{X}_{(1)} - \underline{\mu}_{(1)}))\,(\underline{X}_{(2)} - \underline{\mu}_{(2)} - \Sigma_{(21)}\Sigma^{-}_{(11)}\,(\underline{X}_{(11)} - \underline{\mu}_{(1)}))'$$

$$= \Sigma_{(22)} + \Sigma_{(21)}\Sigma^{-}_{(11)}\Sigma_{(11)}\,\Sigma^{-1}_{(11)}\,\Sigma_{(12)} - 2\Sigma_{(21)}\Sigma^{-}_{(11)}\,\Sigma_{(12)}$$

$$= \Sigma_{(22)} - \Sigma_{(21)}\Sigma^{-}_{(11)}\Sigma_{(12)},$$

and

$$E(\underline{X}_{(2)} - \underline{\mu}_{(2)} - \Sigma_{(21)}\Sigma^{-}_{(11)}\,(\underline{X}_{(1)} - \underline{\mu}_{(1)})\,(\underline{X}_{(1)} - \underline{\mu}_{(1)})'$$

$$= \Sigma_{(21)} - \Sigma_{(21)}\,\Sigma^{-}_{(11)}\,\Sigma_{(11)}$$

$$= 0\,. \tag{2.15}$$

From (2.14), (2.15) and Theorems 2.4 and 2.5 we conclude that

$\underline{X}_{(2)} - \underline{\mu}_{(2)} - \Sigma_{(21)}\,\Sigma^{-}_{(11)}\,(\underline{X}_{(1)} - \underline{\mu}_{(1)})$, and $\underline{X}_{(1)} - \underline{\mu}_{(1)}$ are independent, and the conditional distribution of $\underline{X}_{(2)}$ given $\underline{X}_{(1)} = \underline{x}_{(1)}$ is (N-q)-variate normal with mean $\underline{\mu}_{(2)} + \Sigma_{(21)}\,\Sigma^{-}_{(11)}\,(\underline{x}_{(1)} - \underline{\mu}_{(1)})$ and covariance matrix $\Sigma_{(22)} - \Sigma_{(21)}\Sigma^{-}_{(11)}\,\Sigma_{(12)}$.

Q.E.D.

In the above theorems, if Σ is positive definite, then the generalized inverse becomes the true inverse. In such cases the definition always guarantees the existence of the probability density function as given in (2.10) and readers may find it instructive to prove Theorems (2.4), (2.5) and (2.6) with the help of the probability density function. In the development which follows we shall assume Σ to be positive definite as it will be of direct use for analysis of variance.

2.3.1. Distribution of Quadratic Forms

Theorem 2.7. Let $\underline{X} = (X_1,\ldots,X_N)'$ have an N-variate normal distribution with mean $\underline{\mu}$ and positive definite covariance matrix Σ. Then $\underline{X}'A\underline{X}$ is distributed as noncentral chisquare with k degrees of freedom if and only if ΣA is an idempotent matrix of rank k.

Note. If ΣA is an idempotent matrix then $\Sigma A\,\Sigma A = \Sigma A$. Since Σ is positive definite we can equivalently express the stated condition of the theorem as $A\,\Sigma\,A = A$.

Proof. Since Σ is positive definite there exists a non-singular matrix C such that $\Sigma = CC'$. Write $\underline{X} = C\underline{Y}$. From Theorem 2.5, $\underline{Y}$ has an N-variate normal distribution with mean $\underline{\nu} = C^{-1}\underline{\mu}$ and covariance matrix I. Since $\underline{X}'A\underline{X} = \underline{Y}'B\underline{Y}$, where $B = C'AC$, the theorem will be proved if we can establish that $\underline{Y}'B\underline{Y}$ is distributed as a noncentral chisquare with k degrees of freedom if and only if B is an idempotent matrix of rank k. To prove it let us first assume that B is an idempotent matrix of rank k. There exists an orthogonal matrix θ such that $\theta B\theta'$ is a diagonal matrix

$$D = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

where $\underline{I}$ is the identity matrix of dimension k × k. Let $\underline{Z} = (Z_1,\ldots,Z_N)' = \theta\underline{Y}$. Then

$$\underline{Y}'B\underline{Y} = \underline{Z}'D\underline{Z} = \sum_{i=1}^{k} Z_i^2 .$$

Since θ is an orthogonal matrix and the components of $\underline{Y}$ are independently distributed with unit variance, $Z_1,\ldots,Z_k$ are independently normally distributed with unit variance. Hence $\sum_{i=1}^{k} Z_i^2 = \underline{Z}'D\underline{Z}$, and equivalently $\underline{Y}'B\underline{Y}$ has the noncentral chisquare distribution with noncentrality parameter $(\theta\underline{\nu})'D(\theta\nu)$ $=\nu'\ \theta'\underline{D}\theta\ \underline{\nu} = \underline{\nu}'\ B\ \underline{\nu}$.

To prove that it is also a necessary condition let us assume that $\underline{Y}'B\underline{Y}$ is distributed as $\chi_k^2(\underline{\nu}'B\underline{\nu})$. If B is of rank m then there exist an orthogonal matrix θ such that $\theta B\theta'$ is a diagonal matrix with nonzero diagonal elements $\lambda_1,\ldots,\lambda_m$, the characteristic roots of B. We can assume without any loss of generality that the first m diagonal elements are nonzero. Writing $\underline{Z} = \theta\underline{Y}$,

$$\underline{Y}'B\underline{Y} = \sum_{i=1}^{m} \lambda_i Z_i^2.$$

Since Z_i^2, $i = 1,\ldots,m$ are independently distributed as noncentral chisquare random variables each with one degree of freedom and $\underline{Y}'B\underline{Y}$ is distributed as $\chi_k^2(\underline{\nu}'B\underline{\nu})$, it follows from the uniqueness of the characteristic function that $m = k$ and $\lambda_i = 1$, $i = 1,\ldots,m$. In other words $\theta B\theta'$ is a diagonal matrix with the first k diagonal elements equal to unity and the rest are zeros. This in turn, implies that B is idempotent.

Q.E.D.

Theorem 2.8. Let $\underline{X}$ have the N-variate normal distribution with mean $\underline{\mu}$ and positive definite covariance matrix Σ and let

$$\underline{X}'\Sigma^{-1}\underline{X} = \underline{X}'A_1\underline{X} + \underline{X}'A_2\underline{X}, \tag{2.16}$$

where $\underline{X}'A_1\underline{X}$ is distributed as noncentral chisquare $\chi_k^2(\lambda_1^2)$ with k degrees of freedom and the noncentrality parameter λ_1^2. Then $\underline{X}'A_2\underline{X}$ is distributed as noncentral chisquare with N-k degrees of freedom and the noncentrality parameter $\underline{\mu}'\Sigma^{-1}\underline{\mu} - \lambda_1^2$.

Proof. Let C be a nonsingular matrix such that $\Sigma = CC'$ and let $\underline{Y} = C^{-1}\underline{X}$. Obviously $\underline{Y}$ has the N-variate normal distribution with mean $\underline{\nu} = C^{-1}\underline{\mu}$ and I as the covariance matrix. Now

$$\underline{Y}'\underline{Y} = \underline{Y}'B_1\underline{Y} + \underline{Y}'B_2\underline{Y},$$

where $B_i = C'A_iC$, $i = 1, 2$. Since C is nonsingular

$\text{rank}\,(A_i) = \text{rank}(B_i) \quad i = 1, 2.$

The theorem will be proved if we can show that $\underline{Y}'B_2\underline{Y}$ is distributed as the noncentral $\chi^2_{N-k}(\underline{\mu}'\,\Sigma^{-1}\underline{\mu} - \lambda_1^2)$. Since $\underline{Y}'B_1\underline{Y}$ has the noncentral chisquare $\chi^2_k(\lambda_1^2)$, by the above theorem we conclude that B_1 is an idempotent matrix of rank k. But

$$\underline{Y}'B_2\underline{Y} = \underline{Y}'(I-B_1)\underline{Y},$$

and

$$(I-B_1)(I-B_1) = I-2B_1 + B_1 = I-B_1, \tag{2.17}$$

which implies that $I-B_1$ is an idempotent matrix of rank $N-k$. Hence $\underline{Y}'B_2\underline{Y}$ has the noncentral chisquare distribution with $N-k$ degrees of freedom and with the noncentrality parameter

$$\begin{aligned}\underline{\nu}'(I-B_1)\underline{\nu} &= \underline{\nu}'\underline{\nu} - \underline{\nu}'B_1\underline{\nu}\\ &= \underline{\mu}'\,\Sigma^{-1}\underline{\mu} - \underline{\mu}'C'^{-1}C'A_1CC^{-1}\underline{\mu}\\ &= \underline{\mu}'\,\Sigma^{-1}\underline{\mu} - \underline{\mu}'A_1\underline{\mu}\,.\end{aligned}$$

Q.E.D.

Theorem 2.9. Fisher-Cochran's Theorem. Let $\underline{X}$ have the N-dimensional normal distribution with mean $\underline{\mu}$ and positive definite covariance matrix Σ and let

$$\underline{X}'\,\Sigma^{-1}\underline{X} = Q_1 + \ldots + Q_k, \tag{2.18}$$

where $Q_i = \underline{X}'A_i\underline{X}$, with $\text{rank}(A_i) = N_i$, $i = 1,\ldots,k$. Then Q_i's are independently distributed noncentral chisquares $\chi^2_{N_i}(\underline{\mu}'A_i\underline{\mu})$ if and only if $\sum_{i=1}^{k} N_i=N$, in which case $\underline{\mu}'\Sigma^{-1}\underline{\mu} = \sum_{i=1}^{k} \mu'A_i\,\underline{\mu}$.

Proof. Since Σ is positive definite there exists a nonsingular matrix $\underline{C}$ such that $\Sigma = CC'$. Let $\underline{Y} = C^{-1}\underline{X}$. Then (2.18) reduces to

$$\underline{Y}'\underline{Y} = \sum_{i=1}^{k} \underline{Y}'B_i\underline{Y} \tag{2.19}$$

where $B_i = C'A_iC$, and $\underline{Y}$ has an N-dimensional normal distribution with mean $\underline{\nu} = C^{-1}\underline{\mu}$ and with I as the covariance matrix.

Let us suppose that $\underline{Y}'B_i\underline{Y}$, $i = 1,\ldots,k$ are independently distributed as $\chi^2_{N_i}$ $(\underline{\nu}'B_i\underline{\nu})$. Then $\sum_{i=1}^{k} \underline{Y}'B_i\underline{Y}$ is distributed as a noncentral chisquare with $\sum_{i=1}^{k} N_i$ degrees of freedom and the noncentrality parameter $\sum_{i=1}^{k} \underline{\nu}'B_i\underline{\nu}$. Since $\underline{Y}'\underline{Y}$ is distributed as $\chi^2_N(\underline{\nu}'\underline{\nu})$ and (2.19) holds, we conclude from the uniqueness of the characteristic function that $\sum_{i=1}^{N} N_i = N$ and $\sum_{i=1}^{k} \underline{\nu}'B_i\underline{\nu} = \underline{\nu}'\underline{\nu}$, that is, $\sum_{i=1}^{k} \underline{\mu}'A_i\underline{\mu} = \underline{\mu}'\,\Sigma^{-1}\underline{\mu}$.

To prove that these conditions are sufficient let us assume that $\sum_{i=1}^{k} N_i = N$. Since $\underline{Y}'B_i\underline{Y}$ is a quadratic form in $\underline{Y}$ with $\operatorname{rank}(B_i) = N_i$, using the arguments of Theorem 2.7 we can write

$$\underline{Y}'B_i\underline{Y} = \sum_{j=1}^{N_i} \pm Z_{ij}^2 \;, \quad i = 1,\ldots,k, \tag{2.20}$$

where Z_{ij}'s are linear combinations of $Y_1,\ldots,Y_N$. Let $\underline{Z} = (Z_{11},\ldots,Z_{1N_1},\ Z_{k1},\ Z_{kN_k})'$. Then from (2.20) we get

$$\underline{Y}'\underline{Y} = \underline{Z}'\ \Delta\ \underline{Z} \tag{2.21}$$

where Δ is a diagonal matrix of dimension $N \times N$ with diagonal elements $+1$ or -1. Let $\underline{Z} = A\underline{Y}$ be the linear transformation which transforms the positive definite quadratic form $\underline{Y}'\underline{Y}$ to $\underline{Z}'\Delta\underline{Z}$. Since for all values of the vector $\underline{Y}$

$$\underline{Y}'\underline{Y} = \underline{Z}'\Delta\ \underline{Z} = \underline{Y}'A'\ \Delta\ A\underline{Y},$$

we conclude that $A'\Delta\ A = I$. In other words A is a nonsingular matrix of dimension $N \times N$. This implies that $\underline{Z}'\ \Delta\ \underline{Z}$ is a positive definite quadratic form and hence $\Delta = I$, $A'A = I$. Thus A is an orthogonal matrix and hence $\underline{Z}$ has the N-variate normal distribution with I as the covariance matrix. Hence from (2.20) we conclude that $\underline{Y}'B_i\underline{Y}$, $i = 1,\ldots,k$, are independently distributed $\chi^2_{N_i}(\underline{\nu}'B_i\underline{\nu})$. But $\underline{Y}'\underline{Y}$ is distributed as $\chi^2_N(\underline{\nu}'\underline{\nu})$. Hence $\underline{\nu}'\underline{\nu} = \sum_{i=1}^{k} \underline{\nu}'B_i\underline{\nu}$, that is, $\underline{\mu}'\,\Sigma^{-1}\underline{\mu} = \sum_{i=1}^{k} \underline{\mu}'A_i\underline{\mu}$.

Q.E.D.

Theorem 2.10. Let $\underline{X}$ have an N-variate normal distribution with mean $\underline{\mu}$ and positive definite covariance matrix Σ, and let the two quadratic forms $\underline{X}'A_1\underline{X}, \underline{X}'A_2\underline{X}$ be distributed as $\chi^2_{N_1}(\underline{\mu}'A_1\underline{\mu})$ and $\chi^2_{N_2}(\underline{\mu}'A_2\underline{\mu})$, respectively. Then $\underline{X}'A_1\underline{X}, \underline{X}'A_2\underline{X}$ are independently distributed if $A_1 \Sigma A_2 = 0$.

Proof. Using the notation of the above theorem we write

$$\underline{X}'A_i\underline{X} = \underline{Y}'B_i\underline{Y}, \quad i = 1,2,$$

where $B_i = C'A_iC$. Since $\underline{Y}'B_i\underline{Y}$, $i = 1,2,$ are distributed as noncentral chisquare with N_i degrees of freedom we conclude from Theorem 2.7 that B_1, B_2 are idempotent matrices of ranks N_1, N_2 respectively. Hence $B_iB_i = B_i$, $i = 1, 2$. Thus

$$B_1(I-B_1-B_2) = (I-B_1-B_2)\, B_2 = 0$$

if $B_1B_2 = 0$. This implies that

$$\text{rank}(B_1) + \text{rank}(B_2) + \text{rank}(I-B_1-B_2) = N.$$

But

$$\underline{Y}'\underline{Y} = \underline{Y}'B_1\underline{Y} + \underline{Y}'B_2\underline{Y} + \underline{Y}'(I-B_1-B_2)\underline{Y} .$$

By the Fisher-Cochran's Theorem the quadratic forms $\underline{Y}'B_1\underline{Y}$, $\underline{Y}'B_2\underline{Y}$, $\underline{Y}'(I-B_1-B_2)\underline{Y}$ are mutually independent.

Note that $B_1B_2 = 0$ implies $A_1\Sigma A_2 = 0$. Q.E.D.

Theorem 2.11. Cochran's Theorem. Let $\underline{X}^\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})'$, $\alpha = 1,\ldots,N$, be a random sample of size N from a p-variate normal distribution with mean $\underline{0}$ and the same positive definite covariance matrix Σ. Assume that

$$\sum_{\alpha=1}^{N} (\underline{X}^\alpha)'\underline{X}^\alpha = Q_1 + \ldots + Q_k , \tag{2.22}$$

where

$$Q_i = \sum_{\alpha,\beta=1}^{N} (\underline{X}^\alpha)'\, a^i_{\alpha\beta}\underline{X}^\beta, \quad i = 1,\ldots,k ,$$

and $A_i = (a^i_{\alpha\beta})$ is of rank N_i. Then Q_i's are independently distributed as

$$\sum_{\alpha=N_1+\dots+N_{i-1}+1}^{N_1+\dots+N_i} (\underline{Z}^{\alpha})'(\underline{Z}^{\alpha}) \tag{2.23}$$

where $\underline{Z}^{\alpha} = (Z_{\alpha 1}, \dots, Z_{\alpha p})'$, $\alpha = 1, \dots, \sum_{i=1}^{k} N_i$, are independently distributed normal p-vectors with mean $\underline{0}$ and covariance matrix Σ, if and only if $\sum_{i=1}^{k} N_i = N$.

Note. If $p = 1$, this gives the popular univariate version of Cochran's Theorem.

Proof. Let Q_i's be independently distributed as in (2.23). Then $\sum_{i=1}^{k} Q_i$ is distributed as

$$\sum_{i=1}^{k} \sum_{\alpha=1}^{N_i} (\underline{Z}^{\alpha})'\underline{Z}^{\alpha}.$$

Using (2.22) and the uniqueness of the characteristic function we conclude that $\sum_{i=1}^{k} N_i = N$.

To prove the sufficiency of the condition let us assume that $\sum_{i=1}^{k} N_i = N$. Using the argument of the Fisher-Cochran theorem we assert that there exists

$$B = \begin{pmatrix} B_1 \\ \vdots \\ B_k \end{pmatrix}$$

such that $A_i = B_i B_i'$ and B_i is of dimension $N_i \times N$. Let $B = (b_{\alpha\beta})$. Since B is orthogonal,

$$\underline{Z}^{\alpha} = \sum_{\beta=1}^{N} b_{\alpha\beta}\, \underline{X}^{\beta}, \alpha = 1, \dots, N,$$

are independently distributed p-vectors, with mean $\underline{0}$ and covariance matrix Σ. Clearly

$$Q_i = \sum_{\beta=1}^{N} (\underline{X}^{\alpha})'\, a^{i}_{\alpha\beta}\, \underline{X}^{\beta}$$

$$= \sum_{\alpha=N_1+\dots+N_{i-1}+1}^{N_1+\dots+N_i} (\underline{Z}^{\alpha})'(\underline{Z}^{\alpha})$$

Q.E.D

Example 2.1. Let $X_1,\dots,X_N$ be a random sample of size N from a normal population with mean μ and variance σ^2. Let

$\underline{X} = (X_1,\dots,X_N)'$,

$$A_1 = \begin{pmatrix} 1-\frac{1}{N}, -\frac{1}{N}, \dots, -\frac{1}{N} \\ -\frac{1}{N}, 1-\frac{1}{N}, \dots, -\frac{1}{N} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ -\frac{1}{N}, -\frac{1}{N}, \dots, 1-\frac{1}{N} \end{pmatrix} \quad A_2 = \begin{pmatrix} \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \\ \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \end{pmatrix}$$

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i .$$

Observe that A_1, A_2 are idempotent matrices of rank N-1 and 1, respectively. Furthermore $A_1A_2 = 0$, and

$$\underline{X}'\underline{X} = \sum_{i=1}^{N} (X_i-\bar{X})^2 + N\bar{X}^2$$

$$= \underline{X}'A_1\underline{X} + \underline{X}'A_2\underline{X} .$$

Thus, $\sum_{i=1}^{N} (X_i-\bar{X})^2/\sigma^2$ is distributed independently of $N\bar{X}^2$ as χ^2_{N-1}. Note that $E(\underline{X}'A_1\underline{X}) = 0$.

Example 2.2. Let X_{ij}, $j = 1,\dots,N_i$, $i = 1,\dots,k$, be a random sample of size N_i from the ith normal population with mean μ_i and the same variance σ^2. Let

$$\underline{X} = (X_{11},\dots,X_{1N_1}, X_{21},\dots,X_{2N_2}, \dots, X_{k1},\dots,X_{kN_k})' ,$$

$$x_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij} , \quad X_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} X_{ij} , \quad N = \sum_{i=1}^{k} N_i ,$$

and let the matrices A_1, A_2, A_3, each of dimension $N \times N$, be given by

$$A_3 = \begin{pmatrix} \frac{1}{N}, \dots, \frac{1}{N} \\ \cdot, \dots, \cdot \\ \frac{1}{N}, \dots, \frac{1}{N} \end{pmatrix}$$

$$A_2 = B - A_3, \quad A_1 = I-B ,$$

where B is a diagonal matrix in the block form with diagonal elements $B_{(11)},\dots,B_{(kk)}$, $B_{(ii)}$ being the $N_i \times N_i$ matrix each of whose elements is equal to $1/N_i$ and I is the identity matrix of dimension $N \times N$. It is easy to verify that A_1, A_2, A_3 are all idempotent matrices and rank $(A_1) = N-k$, rank $(A_2) = k-1$, rank $(A_3) = 1$. Furthermore, $A_iA_j = 0$ for $i \neq j$.

Now

$$\underline{X}'\underline{X} = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij}-X_{i.})^2 + \sum_{i=1}^{k} N_i(X_{i.}-X_{..})^2 + N X_{..}^2$$

$$= \underline{X}'A_1\underline{X} + \underline{X}'A_2\underline{X} + \underline{X}'A_3\underline{X}.$$

Noting that

$$E(\underline{X}'A_1\underline{X}) = 0, \quad E(\underline{X}'A_2\underline{X}) = \sum_{i=1}^{k} N_i(\mu_i-\mu_.)^2,$$

$$E(\underline{X}'A_3\underline{X}) = N\mu_.^2,$$

where $\mu_. = \sum_{i=1}^{k} N_i\mu_i / N$, we conclude that $\underline{X}'A_1\underline{X}$ and $\underline{X}'A_2\underline{X}$ are independently distributed as χ^2_{N-k} and $\chi^2_{k-1}\left(\sum_{i=1}^{k} N_i(\mu_i-\mu_.)^2\right)$, respectively.

2.3.2. *Wishart Distribution and Generalized Variance*: Let $\underline{X}^\alpha = (X_{\alpha 1},\dots,X_{\alpha p})'$, $\alpha = 1,\dots,N$, be a random sample of size N from a p-variate normal distribution with mean $\underline{\mu}$ and positive definite covariance matrix Σ and let

$$S = \sum_{\alpha=1}^{N}(\underline{X}^\alpha-\bar{\underline{X}})(\underline{X}^\alpha-\bar{\underline{X}})', \quad \bar{\underline{X}} = \frac{1}{N}\sum_{\alpha=1}^{N}\underline{X}^\alpha. \tag{2.24}$$

For given sample values $\underline{x}^\alpha$, $\alpha = 1,\dots,N$,

$$\frac{s}{N} = \frac{1}{N}\sum_{\alpha=1}^{N}(\underline{x}^\alpha-\bar{\underline{x}})(\underline{x}^\alpha-\bar{\underline{x}})', \quad \bar{\underline{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\underline{x}^\alpha$$

are the maximum likelihood estimates of Σ, $\underline{\mu}$ respectively.

Theorem 2.12. (i) $\sqrt{N}\,\bar{\underline{X}}$ is distributed independently of S as p-variate normal with mean $\sqrt{N}\,\underline{\mu}$ and positive definite covariance matrix Σ and S is distributed as

$$\sum_{\alpha=1}^{N-1}(\underline{Z}^\alpha)(\underline{Z}^\alpha)'$$

where $\underline{Z}^{\alpha}$, $\alpha = 1,\ldots,N-1$ are independently distributed p-variate normal vectors with the same mean $\underline{0}$ and the same covariance matrix Σ ;

(ii) S is positive definite with probability 1 if and only if $N > p$.

Proof. (i) Let θ be an $N \times N$ orthogonal matrix of the following form

$$\theta = \begin{pmatrix} \frac{1}{\sqrt{N}}, & \ldots, & \frac{1}{\sqrt{N}} \\ \theta_{21}, & \ldots, & \theta_{2N} \\ \cdot\,, & \ldots, & \cdot \\ \theta_{N1}, & \ldots, & \theta_{NN} \end{pmatrix}$$

and let

$$\underline{Z}^{\alpha} = \sum_{\beta=1}^{N} \theta_{\alpha\beta}\underline{X}^{\beta}, \quad \alpha = 1,\ldots, N,$$

with $\theta_{1\beta} = 1/\sqrt{N}$ for all β . The set of random vectors $\underline{Z}^{\alpha}$ has a joint normal distribution because the entire set of components is a set of linear combinations of the set of the normal random vectors $\underline{X}^{\alpha}$, $\alpha = 1,\ldots,N$. Now

$$E(\underline{Z}^{\alpha}) = \begin{cases} \sqrt{N}\,\underline{\mu}, & \text{if } \alpha = 1 \\ \sum_{\beta=1}^{N} \theta_{\alpha\beta}\underline{\mu} = \sqrt{N}\underline{\mu} \sum_{\beta=1}^{N} \theta_{\alpha\beta}\frac{1}{\sqrt{N}} = 0, & \text{if } \alpha > 1, \end{cases}$$

$$\begin{aligned} \mathrm{Cov}(\underline{Z}^{\alpha},\underline{Z}^{\gamma}) &= \sum_{\beta=1}^{N} \theta_{\alpha\beta}\theta_{\gamma\beta}E(\underline{X}^{\beta}-\underline{\mu})(\underline{X}^{\gamma}-\underline{\mu})' \\ &= \begin{cases} 0, & \text{if } \alpha \neq \gamma \\ \Sigma, & \text{if } \alpha = \gamma . \end{cases} \end{aligned}$$

Thus $\underline{Z}^{\alpha}$, $\alpha = 1,\ldots,N$ are independently distributed normal p-vectors with means and covariance matrices as given above. Since θ is orthogonal

$$\sum_{\alpha=1}^{N} \underline{Z}^{\alpha}(\underline{Z}^{\alpha})' = \sum_{\alpha=1}^{N} \underline{X}^{\alpha}(\underline{X}^{\alpha})' .$$

Thus

$$S = \sum_{\alpha=2}^{N} \underline{Z}^{\alpha}(\underline{Z}^{\alpha})', \quad \sqrt{N}\,\underline{\bar{X}} = \underline{Z}_1$$

are independently distributed.

(ii) Let

$$B = (\underline{Z}^2, \ldots, \underline{Z}^N).$$

Then $S = BB'$. Since B is a random matrix of dimension $p \times (N-1)$, this part will be proved if we can show that B has rank p if and only if $N > p$. Obviously by adding more columns to B we cannot diminish its rank and if $N \le p$, then the rank of B is less than p.

Thus it will be sufficient to show that B has rank p with probability one when $N-1 = p$. For any set of $(p-1)$ p-vectors $\underline{\alpha}^1, \ldots, \underline{\alpha}^{p-1}$ in E^p let $S(\underline{\alpha}^1, \ldots, \alpha^{p-1})$ denote the subspace spanned by $\underline{\alpha}^1, \ldots, \alpha^{p-1}$. Since the covariance matrix of each $\underline{Z}^{\alpha}$ is nonsingular, for any $\underline{\alpha}^1, \ldots, \underline{\alpha}^{p-1}$ in E^p.

$\text{Prob}\ \{\underline{Z}^{\alpha} \in S(\underline{\alpha}^1, \ldots, \underline{\alpha}^{p-1})\} = 0$ for $\alpha = 1, \ldots, p$. Since $\underline{Z}^1, \ldots, \underline{Z}^p$ are independent and identically distributed

$$\text{Prob}\ \{\underline{Z}^1, \ldots, \underline{Z}^p \text{ are linearly dependent}\}$$

$$\le \sum_{i=1}^{p} \{\text{Prob}\ \underline{Z}^i \in S(\underline{Z}^1, \ldots, \underline{Z}^{i-1}, \underline{Z}^{i+1}, \ldots, \underline{Z}^p)\}$$

$$= p\ \text{Prob}\ \{\underline{Z}^1 \in S(\underline{Z}^2, \ldots, \underline{Z}^p)\}$$

$$= p\ E\ [\text{Prob}\ \{\underline{Z}^1 \in S(\underline{z}^2, \ldots, \underline{z}^p) \mid \underline{Z}^2 = \underline{z}^2, \ldots, \underline{Z}^p = \underline{z}^p\}]$$

$$= p\ E(0) = 0.$$

Q.E.D.

The distribution of S is known as the Wishart distribution with N-1 degrees of freedom and with parameter Σ. Its probability density function $W_p(N-1, \Sigma)$ (see Giri (1976), for a proof) is

$$W_p(N-1,\Sigma) = \begin{cases} K\ |\Sigma|^{-\frac{1}{2}(N-1)}\ |s|^{\frac{1}{2}(N-p-2)}\ \exp\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}s\}\ , \\ \qquad \text{if } s \text{ is positive definite,} \\ 0\ , \quad \text{otherwise,} \end{cases} \tag{2.25}$$

where

$$K^{-1} = \prod_{i=o}^{p-1} \Gamma(\tfrac{1}{2}(N-1-i))\ 2^{\frac{1}{2}(N-1)p}\ \pi^{p(p-1)/4} \tag{2.26}$$

If $p = 1$, Σ is a scalar quantity and S is a scalar random variable, and $W_p(N-1,\Sigma)$ reduces to the probability density function of Σ times a central chisquare random variable with $N-1$ degrees of freedom. The Wishart distribution is thus the multivariate generalization of a central chisquare distribution and it plays the same central role as that of chisquare in multivariate analysis of variance and covariance.

In univariate analysis of variance, the F-statistic which is related to the ratio of two independent chisquare random variables plays a dominant role for testing various hypotheses in linear models. In the multivariate case, the ratio of the determinants of two independent Wishart matrices plays a similar dominant role.

The characteristic function of S is given by

$$E(\exp\{i\ \mathrm{tr}\ \theta S\}) = |I-2i\Sigma\theta|^{-\frac{1}{2}(N-1)}\ , \tag{2.27}$$

where θ is a real symmetric matrix of dimension $p \times p$. From (2.27) it follows that $E(S) = (N-1)\Sigma$. Thus s/N is not an unbiased estimate of Σ and the unbiased estimate of Σ is $s/(N-1)$. Furthermore, if $S_1,\ldots,S_k$ are independently distributed Wishart (random) matrices with N_i, $i = 1,\ldots,k$, degrees of freedom and the same parameter Σ, then $\sum_{i=1}^{k} S_i$ is distributed as Wishart with $\sum_{i=1}^{k} N_i$ degrees of freedom and the parameter Σ.

For a p-variate normal population with covariance matrix Σ, Wilks (1932) called $|\Sigma|$ the generalized variance of the population. For $p = 1$ it reduces to the ordinary variance and for $p \geq 1$ it is a measure of the internal scatter in the population. In applications it is often used to rank several distinct multivariate populations in order of their internal scatter. Similarly the generalized variance of a sample $\underline{x}^{\alpha}, \alpha = 1,\ldots,N$, from the population is given by

$$\left|\frac{s}{N-1}\right| = \left|\frac{\sum_{\alpha=1}^{N} (\underline{x}^{\alpha}-\bar{\underline{x}})(\underline{x}^{\alpha} - \bar{\underline{x}})'}{N-1}\right|$$

where $\bar{\underline{x}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underline{x}^{\alpha}$.

Theorem 2.13. Let S be distributed as $W_p(N-1,\Sigma)$. Then $|S|$ is distributed as $|\Sigma| \prod_{i=1}^{p} \chi^2_{N-i}$, where χ^2_{N-i}, $i = 1,\ldots,p$, are independent central chisquare random variables.

Proof Since Σ is positive definite there exists a nonsingular matrix C such that $C \Sigma C' = I$. Write $S^* = CSC'$. Since S is distributed as $W_p(N-1,\Sigma)$ we can write S as

$$S = \sum_{\alpha=1}^{N-1} \underline{Z}^{\alpha}(\underline{Z}^{\alpha})',$$

where $\underline{Z}^{\alpha}$, $\alpha = 1,\ldots,N-1$ are independent normal p-vectors with mean $\underline{0}$ and covariance matrix Σ. Now

$$S^* = \sum_{\alpha=1}^{N-1} (C\underline{Z}^{\alpha})(C\underline{Z}^{\alpha})' = \sum_{\alpha=1}^{N-1} \underline{Y}^{\alpha}(\underline{Y}^{\alpha})'$$

where $\underline{Y}^{\alpha} = C\underline{Z}^{\alpha}$, $\alpha = 1,\ldots,N-1$, are independent normal p-vectors with mean $\underline{0}$ and covariance matrix I. Thus S^* is distributed as $W_p(N-1,I)$ and its probability density function (from 2.25) is

$$K|s^*|^{\frac{1}{2}(N-2-p)} \exp\{-\tfrac{1}{2}\mathrm{tr}\, s^*\}. \tag{2.28}$$

Partition $S^* = (S^*_{ij})$ as

$$S^* = \begin{pmatrix} S^*_{(11)}, & S^*_{(12)} \\ S^*_{(21)}, & S^*_{pp} \end{pmatrix}.$$

Since S^* is symmetric and $|s^*| = s^*_{pp}|s^*_{(11)} - s^*_{(12)} s^{*-1}_{pp} s^*_{(21)}|$, the joint probability density function of $S^*_{(11)}, S^*_{(12)}, S^*_{pp}$ is

$$K(s^*_{pp})^{\frac12(N-p-2)} \exp\{-\tfrac12 s^*_{pp}\}$$

$$\times |s^*_{(11)} - s^*_{(12)} s^{*-1}_{pp} s^*_{(21)}|^{\frac12(N-p-2)} \exp\{-\tfrac12 \mathrm{tr}(s^*_{(11)} - s^*_{(12)} s^{*-1}_{pp} s^*_{(21)})\}$$

$$\times \exp\{-\tfrac12 (s^*_{(12)} s^{*-1}_{pp} s^*_{(21)})\}.$$

Let

$$W = S^*_{(11)} - S^*_{(12)} S^{*-1}_{pp} S^*_{(21)},$$

$$U = S^*_{(12)} S^{*-\frac12}_{pp}, \quad (S^{*-\frac12}_{pp} \text{ being the positive square root of } S^{*-1}_{pp})$$

$$V = S^*_{pp} .$$

The Jacobian of the transformation $(S^*_{(11)}, S^*_{(12)}, S^*_{pp}) \to (W,U,V)$ is given by the absolute value of the determinant of the following matrix of partials .

$$\begin{array}{c} \\ s^*_{(11)} \\ s^*_{(12)} \\ s^*_{pp} \end{array} \begin{pmatrix} w & u & v \\ I & - & - \\ 0 & A & - \\ 0 & 0 & I \end{pmatrix}$$

where "-" indicates some matrix which need not be evaluated specifically and A is the matrix of partial derivatives of the transformation

$$s^*_{(12)} \to u\, v^{\frac12} \quad (v \text{ fixed}) . \tag{2.29}$$

Since the Jacobian of the transformation (2.29) is $|v|^{\frac12(p-1)}$, the joint probability density function of W,U,V is given by,

$$K\,(v)^{\frac12(N-2)} \exp\{-\tfrac12 v\}$$

$$\times |w|^{\frac12(N-2-(p-1)-1)} \{\exp -\tfrac12 \mathrm{tr}'\, w\}$$

$$\times \exp\{-\tfrac12 \mathrm{tr}\; uu'\}. \tag{2.30}$$

From this we conclude that S^*_{pp} is distributed as $W_1(N-1)$ independently of $W = S^*_{(11)} - S^*_{(12)} S^{*-1}_{pp} S^*_{(21)}$ and W is distributed as $W_{p-1}(N-2, I_{p-1})$ where I_{p-1} is the identity matrix of dimension $(p-1)\times(p-1)$. Noting that $W_1(N-1,1)$ is a central chisquare with $N-1$ degrees of freedom we conclude that $|W_p(N-1,I_p)|$ is distributed as the product of χ^2_{N-1} and $|W_{p-1}(N-2,I_{p-1})|$, where χ^2_{N-1}, $W_{p-1}(N-2,I_{p-1})$ are independent. Repeating the above argument $p-1$ times we conclude that $|S^*|$ is distributed as $\Sigma_{i=1}^{p}\chi^2_{N-i}$, where χ^2_{N-i}, $i = 1,\ldots,p$, are independent.

Q.E.D.

Hotelling's T^2-statistic, a multivariate extension of the square of the Student-t statistic, is fundamental in testing the hypothesis of equality of the mean vectors of multivariate normal populations with unknown but equal covariance matrix and plays the same role in the multivariate case as the Student-t in the univariate case. We shall now discuss the distribution of T^2-statistic in the general setup. Let $\underline{X}^{\alpha} = (X_{\alpha 1},\ldots,X_{\alpha p})'$, $\alpha = 1,\ldots,N$ be a random sample of size N from a normal population with mean $\underline{\mu}$ and positive definite covariance matrix Σ. Let

$$T^2 = N(N-1)\ \underline{\bar{X}}'\underline{S}^{-1}\underline{\bar{X}}\ ,$$

where $\underline{\bar{X}},\underline{S}$ are as defined earlier.

Theorem 2.14. $T^2/(N-1)$ is distributed as $\chi^2_p(N\ \underline{\mu}'\ \Sigma^{-1}\ \underline{\mu})/\chi^2_{N-p}$ where $\chi^2_p(.)$ and χ^2_{N-p} are independent.

Proof. Let C be a nonsingular matrix such that $C\Sigma C'=I$ and let $\underline{Y} = \sqrt{N}\ C\ \underline{\bar{X}}$, $A = CSC'$ and $\underline{\nu} = \sqrt{N}C\underline{\mu}$. Then $\underline{Y}$ has a p-variate normal distribution with mean $\underline{\nu}$ and covariance matrix I and A is distributed as $\Sigma_{\alpha=1}^{N-1}\underline{Y}^{\alpha}(\underline{Y}^{\alpha})'$, where $\underline{Y}^{\alpha}$, $\alpha = 1,\ldots,N-1$ are independent normal p-vectors with mean $\underline{0}$ and covariance matrix I.

Consider a random orthogonal matrix Q of dimension $p\times p$ whose first row is $\underline{Y}'(\underline{Y}'\underline{Y})^{-\frac{1}{2}}$ and the other $p-1$ rows are defined arbitrarily. Let

$$\begin{aligned} U &= (U_1,\ldots,U_p)' \\ &= Q\ \underline{Y}\ , \end{aligned}$$

$$B = (B_{ij}) = QAQ' .$$

It is easy to check that

$$U_1 = (\underline{Y}'\underline{Y})^{\frac{1}{2}} ,$$

$$U_i = 0, \; i = 2,\ldots, p,$$

and

$$\begin{aligned} N\,\underline{\bar{X}}'\, S^{-1}\, \underline{\bar{X}} &= \underline{Y}'\, A^{-1}\, \underline{Y} \\ &= \underline{U}'\, B^{-1}\, \underline{U} \\ &= U_1^2/(B_{11} - B_{(12)}B_{(22)}^{-1}B_{(21)}) , \end{aligned}$$

where B is partitioned as

$$B = \begin{pmatrix} B_{11}, & B_{(12)} \\ B_{(21)}, & B_{(22)} \end{pmatrix} .$$

Since the conditional distribution of B given Q is Wishart with N-1 degrees of freedom and the parameter I, by computation similar to (2.30) we conclude that the conditional distribution of $B_{11}-B_{(12)}B_{(22)}^{-1}B_{(21)}$ given Q is the central chisquare with N-p degrees of freedom (Wishart $W_1(N-p, 1)$). As this conditional distribution does not depend on Q the unconditional distribution of $B_{11}-B_{(12)}B_{(22)}^{-1}B_{(21)}$ is also a central chisquare with N-p degrees of freedom. By Theorem 2.7, $\underline{Y}'\underline{Y}$ is distributed as a noncentral chisquare with p degrees of freedom and the noncentrality parameter $\underline{\nu}'\underline{\nu} = N\underline{\mu}'\,\Sigma^{-1}\,\underline{\mu}$. The independence of $\underline{Y}'\underline{Y}$ and $B_{11}-B_{(12)}B_{(22)}^{-1}B_{(21)}$ is clear. Hence the theorem.

Q.E.D.

2.3.3. Complex Multivariate Normal Distribution: A complex random variable $Z = X + iY$, $i = \sqrt{-1}$ has a complex normal distribution if its real and imaginary parts have a joint bivariate normal distribution. A p-variate complex random vector $\underline{Z} = (Z_1,\ldots,Z_p)'$ with $Z_j = X_j + iY_j$, $j = 1,\ldots,p$, has a normal distribution if the real 2p-vector $(X_1,\ldots,X_p,Y_1,\ldots,Y_p)'$ has a 2p-variate normal distribution. Let $\underline{Z}$ be a complex vector with $E(\underline{Z}) = \underline{\alpha}$ and Hermitian complex covariance matrix $E(\underline{Z} - \underline{\alpha})(\underline{Z} - \underline{\alpha})^* = \Sigma$, where $(\underline{Z} - \underline{\alpha})^*$ is the adjoint of $(\underline{Z} - \underline{\alpha})$. Its probability density function is given by

$$f_{\underline{Z}}(\underline{z}) = (2\pi)^{-p}|\Sigma|^{-1} \exp\{-(\underline{z}-\underline{\alpha})^* \Sigma^{-1}(\underline{z}-\underline{\alpha})\}. \tag{2.31}$$

The special case with the added restriction that

$$E(\underline{Z} - \underline{\alpha})(\underline{Z} - \underline{\alpha})' = 0$$

is of considerable interest in practical applications. This condition implies that the real and imaginary parts of different components are pairwise independent and the real and the imaginary parts of the same components are independent with the same variance. With the above probability density function one can establish in a straightforward way that, as in the real case, if $\underline{Z}$ has a complex p-variate normal distribution with mean $\underline{\alpha}$ and Hermitian positive definite covariance matrix Σ, then $C\underline{Z}$, where C is a complex nonsingular matrix of dimension $p \times p$, has a complex p-variate normal distribution with mean $C\underline{\alpha}$ and Hermitian positive definite covariance matrix $C \Sigma C^*$, and any subset of the components of $\underline{Z}$ has a complex multivariate normal distribution. Furthermore if the covariance between any two subsets of the components of $\underline{Z}$ is 0 then these subsets are stochastically independent. Other results related to analysis of variance (complex case) will be included in the exercise as problems.

2.3.4. Multinomial Distribution: For the analysis of variance of discrete data in the form of observed frequencies in discrete categories we need to outline the multinomial distribution. Let $\underline{X} = (X_1,\dots,X_p)'$ be a discrete p-dimensional random vector with probability mass function

$$p_{X_1,\dots,X_p}(x_1,\dots,x_p) = \begin{cases} \dfrac{n!}{x_1!\dots x_p!} \prod_{i=1}^{p} p_i^{x_i}, & \text{if } 0 \le x_i \le n \text{ for all } i, \\ 0, & \text{otherwise,} \end{cases} \tag{2.32}$$

where $p_i \ge 0$, subject to the restriction that $\sum_{i=1}^{p} p_i = 1$, and $\sum_{i=1}^{p} x_i = n$. It can be verified that

$$E(X_i) = np_i, \quad \operatorname{Var}(X_i) = np_i(1-p_i),$$
$$\operatorname{Cov}(X_i, X_j) = -np_ip_j \quad (i \ne j).$$

The marginal distribution of $(X_1,\ldots,X_q)'$, $q \le p$ is given by

$$p_{X_1,\ldots,X_q}(x_1,\ldots,x_q) = \frac{n!}{x_1!\ldots x_q!(n-n_o)!}\prod_{i=1}^{q} p_i^{x_i}(1-p_1-\ldots-p_q)^{n-n_o}, \tag{2.33}$$

where $\sum_{i=1}^{q} x_i = n_o$.

Let

$$Y_i = \frac{X_i - np_i}{\sqrt{n}}, \quad i = 1,\ldots,p.$$

Then, as $n \to \infty$, $(Y_1,\ldots,Y_{p-1})'$ tends to a multivariate normal distribution.

2.4. INVARIANCE IN STATISTICAL TESTING

Invariance is a mathematical term for symmetry and in practice many statistical testing problems exhibit symmetries with respect to certain transformations in the sample space. It is a generally accepted principle that if a statistical problem with a unique solution is invariant under certain transformations then the solution should be invariant under those transformations.

Example 2.3. Consider the problem of testing the hypothesis that the mean of a univariate normal population with unknown variance σ^2 is equal to a specified constant μ_o. Since $\mu = \mu_o$ implies that for scalar constants $a(\neq 0)$, b, $a\mu + b = a\mu_o + b$, we should expect that our test procedure should depend neither on the scale of measurements nor on the origin of measurements. For example if the measurements are lengths, our test procedures should not differ whether the measurements are in inches or centimeters and whatever may be the origin of measurements.

Let $\mathcal{X}$ be the sample space and A be the σ-algebra of subsets of $\mathcal{X}$ which contains the null set, the whole space and is closed under complementation and countable union. Let Ω the parametric space and let ρ denote the family of probability measures P_θ, $\theta \in \Omega$ on A. We shall be concerned here with the problem of testing the null hypothesis $H_0: \theta \in \Omega_{H_0}$ against the alternatives $H_1 : \theta \varepsilon \Omega_{H_1}$ where Ω_{H_o} and Ω_{H_1} are disjoint subsets of Ω.

The principle of invariance for testing problems involves transformations mainly on the sample space $\mathcal{X}$ and on the parametric space Ω. Among these the most basic is the transformation g on $\mathcal{X}$. The transformation on Ω is the transformation $\bar{g}$ on Ω, induced by g. All transformations considered in the context of invariance will be assumed to have the following properties:

(i) One-to-one from $\mathcal{X}$ onto $\mathcal{X}$, that is, for every $x_1 \varepsilon \mathcal{X}$ there exists $x_2 \varepsilon \mathcal{X}$ such that $x_2 = g(x_1)$ and $g(x_1) = g(x_2)$ implies that $x_1 = x_2$;

(ii) bimeasurable: this is merely to ensure that whenever X with values in $\mathcal{X}$ is a random variable, g(X), usually written as gX is also a random variable with values in $\mathcal{X}$ and for any set $A \varepsilon \mathcal{A}$

$gA = \{gx : x \varepsilon A\}$ and $g^{-1}A$ both belong to $\mathcal{A}$.

The induced transformation $\bar{g}$ on Ω corresponding to the transformation g on $\mathcal{X}$ is defined as follows:

If the random variable X with values in the sample space $\mathcal{X}$ has distribution P_θ, $\theta \varepsilon \Omega$, then gX with values in $\mathcal{X}$ is a random variable with distribution $P_{\theta'}$, where $\theta' = \bar{g}\,\theta \varepsilon \Omega$. This is usually known as the condition of invariance of the family of distributions $\{P_\theta, \theta \varepsilon \Omega\}$.

An equivalent way of stating this is, for every $A \varepsilon \mathcal{A}$

$$P_\theta \{X \varepsilon g^{-1}A\} = P_{\theta'} \{X \varepsilon A\} \tag{2.34}$$

which can also be written as

$$P_\theta \{g^{-1}A\} = P_{\theta'} \{A\} \tag{2.35}$$

where g^{-1} is the inverse transformation corresponding to g and the subscript θ in the left member of (2.34) and (2.35) indicates the distribution of the random variable X and the subscript $\theta' = \bar{g}\theta$ in the right member of these equations indicates the distribution of the random variable gX. It is now evident that (2.35) can also be written as

$$P_\theta\{A\} = P_{\theta'}\{gA\} . \tag{2.36}$$

In terms of mathematical expectation (2.34) or its equivalent can be stated as follows. For any integrable real valued function ϕ

$$E_\theta(\phi(g^{-1}X)) = E_{\bar{g}\theta}(\phi(X)), \tag{2.37}$$

where E_θ denotes the expectation of ϕ when the random variable X has distribution P_θ and $E_{\bar{g}\theta}$ refers to the distribution $P_{\bar{g}\theta}$ of X.

If, in addition, all $P_\theta, \theta \in \Omega$ are distinct, that is, if $\theta_1 \neq \theta_2$, $\theta_1 \theta_2 \in \Omega$, implies that $P_{\theta_1} \neq P_{\theta_2}$ then g determines $\bar{g}$ uniquely. In other words for any θ, $\bar{g}\theta$ is unique. To see this, let for $A \in \mathcal{A}$, let

$$P_\theta(A) = P_{\theta'}(gA) = P_{\theta''}(gA)$$

for $\theta, \theta', \theta''$ belonging to Ω. Since all P_θ, $\theta \in \Omega$ are distinct, $\theta' = \theta''$.

On the other hand if (2.34) holds then the unicity of θ' implies that all P_θ, $\theta \in \Omega$ are distinct.

Invariance of the parametric space Ω. The parametric space Ω remains invariant under a one-to-one transformation g operating from the left on $\mathcal{X}$ onto $\mathcal{X}$ if the induced transformation $\bar{g}$ on Ω satisfies (i) $\bar{g}\theta \in \Omega$ for $\theta \in \Omega$, (ii) for any $\theta' \in \Omega$ there exists a $\theta \in \Omega$ such that $\theta' = \bar{g}\theta$. An equivalent way of stating (i) and (ii) is

$$\bar{g}\Omega = \Omega. \tag{2.38}$$

Given a set S of transformations, each leaving Ω invariant, the following theorem will assert that the set S can be extended to a group G whose members will also leave Ω invariant. The transformations of G will be, therefore, one-to-one and bimeasurable if the member of S are so. This result will be of considerable importance to us as we can use well known group-theoretic results for the development of the invariance principle

Theorem 2.15. Let g_1, g_2 be two transformations operating from the left on $\mathcal{X}$ onto $\mathcal{X}$, leave Ω invariant. Then the transformations g_1g_2, g_1^{-1} defined by

$$g_1g_2(x) = g_1(g_2(x)), \quad g_1^{-1}g_1(x) = x$$

for all $x \in \mathcal{X}$ also leave Ω invariant and satisfy

$$\overline{g_1g_2} = \bar{g}_1 \, \bar{g}_2 , \quad \overline{(g_1^{-1})} = \bar{g}_1^{-1} . \tag{2.39}$$

Proof. If the random variable X with values in $\mathcal{X}$ has distribution $P_\theta, \theta \in \Omega$, then the transformation g_1X has distribution $P_{\bar{g}_1\theta}$, $\bar{g}_1\theta \in \Omega$. Since $\bar{g}_1\theta \in \Omega$ and g_2 leaves Ω invariant, the distribution of $g_2g_1(X) = g_2(g_1(X))$ is

$$P_{\overline{g_2g_1}\theta} = P_{\bar{g}_2\bar{g}_1\theta} ,$$

where $\bar{g}_2\bar{g}_1\theta \in \Omega$ for $\theta \in \Omega$. Thus g_2g_1 leaves Ω invariant and $\overline{g_2g_1} = \bar{g}_2\bar{g}_1$. Similarly (ii) can be proved.

Q.E.D.

Let $\bar{G} = \{\bar{g} : g \in G\}$, where G is the group of transformations operating from the left onto itself. Since for g_1, g_2 in G (2.39) holds, we conclude that $\bar{G}$ is closed under composition and inverse. Furthermore if θ is the identity element of G, $\bar{\theta}$ is the identity element of $\bar{G}$. Thus $\bar{G}$ is a group of transformations on Ω corresponding to G on $\mathcal{X}$ and (2.39) states that the correspondence between G and $\bar{G}$ is a homomorphism. As stated earlier if all P_θ, $\theta \in \Omega$ are distinct, $\bar{g} \in G$ is also one-to-one onto itself.

Let us now consider the problem of testing $H_0 : \theta \in \Omega_{H_0}$ against the alternatives $H_1 : \theta \in \Omega_{H_1}$ where Ω_{H_0}, Ω_{H_1} are disjoint subsets of Ω. Let G be the group of transformations which operates from the left on $\mathcal{X}$ and satisfies (2.34) and (2.38). The problem of testing H_0 against H_1 remains invariant under G if for $g \in G$, $A \in \mathcal{A}$

$$P_\theta(g^{-1}A) = P_{\bar{g}\theta}(A) \tag{2.40}$$

and

$$\Omega_{H_1} = \bar{g}\Omega_{H_1}, \quad \Omega_{H_0} = \bar{g}\,\Omega_{H_0} \,. \tag{2.41}$$

Example 2.4. Let $X_1,\ldots,X_n$ be a random sample of size N from a normal population with mean μ and variance σ^2 with corresponding sample values $x_1,\ldots,x_n$. The sample space χ of $X_1,\ldots,X_n$ is the space of all sample values $(x_1,\ldots,x_n)$. Let

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$

$$\theta = (\mu,\sigma^2) \,.$$

The parametric space is $\Omega = \{\theta = (\mu,\sigma^2),\ -\infty < \mu < \infty,\ \sigma^2 > 0\}$ and let $\Omega_{H_0} = \{(0,\sigma^2),\ \sigma^2 > 0\}$, $\Omega_{H_1} = \{(\mu,\sigma^2):\ \mu \neq 0,\ \sigma^2 > 0\}$. The group G of transformations g which leaves the problem invariant is the multiplicative group of scale changes $x_i \to ax_i$, $i = 1,\ldots,n$, where a is any nonzero real. The induced group $\bar{G}$ on Ω is the group of transformations $\bar{g}$ operating from the left on Ω such that $\bar{g}\theta = (a\mu, a^2\sigma^2)$. Obviously $\bar{g}\Omega=\Omega$ for all $g \in G$. As for any $A \in \mathcal{A}$, $g \in G$, $\underline{x} = (x_1,\ldots,x_n) \in g^{-1}A$ implies that $g\underline{x} \in A$, we have with $y_i = ax_i$, $i = 1,\ldots,n$,

$$P_\theta\{g^{-1}A\} = \int_{g^{-1}A} (2\pi)^{-\frac{1}{2}n}(\sigma^2)^{-\frac{1}{2}n} \exp\{-\tfrac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\}dx_1\ldots dx_n$$

$$= \int_A (2\pi)^{-\frac{1}{2}n}(a^2\sigma^2)^{-\frac{1}{2}n} \exp\{-\tfrac{1}{2}\sum_{i=1}^{n}\frac{(y_i-a\mu)^2}{a^2\sigma^2}\}dy_1 \ldots dy_n$$

$$= P_{\bar{g}\theta}(A) \,.$$

Also $\bar{g}\,\Omega_{H_0} = \Omega_{H_0}$, $\bar{g}\,\Omega_{H_1} - \Omega_{H_1}$.

Example 2.5. Let $X_1,\ldots,X_m$ be a random sample of size m from a normal population with mean μ_1 and variance σ^2 and let $Y_1,\ldots,Y_m$ be a random sample of size n from another independent normal population with mean μ_2 and variance σ^2. The sample space of $(X_1,\ldots,X_m,Y_1,\ldots,Y_n)$ is the space of sample values $(x_1,\ldots,x_m,\ y_1,\ldots,y_n)$. Let

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i,\ y = \frac{1}{n}\sum_{i=1}^{n} y_i,\ s_1^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i-\bar{x})^2,\ s_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2,$$

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i,\ \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,\ S_1^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i-\bar{X})^2,\ S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2,$$

$\theta = (\mu_1, \mu_2, \sigma^2)$.

The parametric space Ω is given by

$$\Omega = \{\theta : -\infty < \mu_i < \infty,\ i = 1, 2,\ \sigma^2 > 0\}.$$

For testing the hypothesis $H_0: \mu_1 = \mu_2 = \mu$ (say) against the alternatives $H_1: \mu_1 \neq \mu_2$, we get,

$$\Omega_{H_0} = \{(\mu, \sigma^2),\ -\infty < \mu < \infty\ \sigma^2 > 0\},$$

$$\Omega_{H_1} = \{(\mu_1, \mu_2, \sigma^2),\ \mu_1 \neq \mu_2,\ \sigma^2 > 0\}.$$

The group G of transformations which leaves the problem invariant is the group of affine transformations (a,b), with $a \neq 0$, which transforms each $x_i \to ax_i + b$ and each $y_i \to ay_i + b$. The induced group $\bar{G}$ is the group of transformations $\bar{g}$ such that $\bar{g}\theta = (a\mu_1 + b,\ a\mu_2 + b,\ a^2\sigma^2)$. As for any $A \in \mathcal{A}$ and $g \in G$, $(x_1,\ldots,x_m, y_1,\ldots,y_n) \in g^{-1}A$ implies that

$$(ax_1 + b,\ldots,ax_m + b,\ ay_1 + b,\ldots,ay_n + b) \in A$$

we have with $u_i = ax_i + b$, $v_i = ay_i + b$,

$$P_\theta(\bar{g}^{-1}A) = \int_{\bar{g}A} (2\pi)^{-\frac{1}{2}(m+n)}(\sigma^2)^{-\frac{1}{2}(m+n)} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=1}^{n}(y_1-\mu_2)^2)\} \prod_{i=1}^{m} dx_i \prod_{i=1}^{n} dy_i$$

$$= \int_{A} (2\pi)^{-\frac{1}{2}(m+n)}(\sigma^2)^{-\frac{1}{2}(m+n)} \exp\{-\frac{1}{2a^2\sigma^2}((\sum_{i=1}^{m}(u_i-a\mu_1-b)^2 + \sum_{i=1}^{n}(v_i-a\mu_2-b)^2)\} \prod_{i=1}^{m} du_i \prod_{i=1}^{n} dv_i$$

$$= P_{\bar{g}\theta}(A)\,.$$

Furthermore $\bar{g}\Omega_{H_0} = \Omega_{H_0}$, $\bar{g}\Omega_{H_1} = \Omega_{H_1}$.

Example 2.6. Let $\underline{X}^{\alpha} = (X_{\alpha 1},\dots,X_{\alpha p})'$, $\alpha = 1,\dots,N$, ($N > p$) be a random sample of size N from a p-variate normal distribution with mean $\underline{\mu} = (\mu_1,\dots,\mu_p)'$ and positive definite covariance matrix Σ. Let $\underline{x}^{\alpha} = (x_{\alpha 1},\dots,x_{\alpha p})'$, $\alpha = 1,\dots,N$, denote the corresponding sample values. Write

$$\bar{\underline{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\underline{x}^{\alpha}, \qquad s = \sum_{\alpha=1}^{N}(\underline{x}^{\alpha}-\bar{\underline{x}})(\underline{x}^{\alpha}-\bar{\underline{x}})',$$

$$\bar{\underline{X}} = \frac{1}{N}\sum_{\alpha=1}^{N}\underline{X}^{\alpha}, \qquad S = \sum_{\alpha=1}^{N}(\underline{X}^{\alpha}-\bar{\underline{X}})(\underline{X}^{\alpha}-\bar{\underline{X}})'$$

The sample space $\mathcal{X}$ is the space of all sample values and the parametric space Ω is the space of all values of $\theta = (\underline{\mu},\Sigma)$. For testing $H_0\colon \underline{\mu} = \underline{0}$ against the alternatives $H_1\colon \underline{\mu} \neq \underline{0}$, we get

$$\Omega_{H_0} = \{(0,\Sigma),\ \Sigma \text{ is positive definite}\},$$

$$\Omega_{H_1} = \{(\underline{\mu},\Sigma),\ \mu\neq 0\ ,\ \Sigma \text{ is positive definite}\}.$$

The group of transformations g which leaves the problem invariant is the full linear group $G_{\ell}(p)$ of nonsingular matrices g which transforms each $\underline{X}^{\alpha}$ to $g\underline{X}^{\alpha}$. The induced group $\bar{G}$ is the group of transformations $\bar{g}$ satisfying $\bar{g}\theta = (g\underline{\mu},\ g\,\Sigma\, g')$. As for any $A \in \mathcal{A}$, $(\underline{x}^1,\dots,\underline{x}^N) \in g^{-1}A$ implies that $(g\underline{x}^1,\dots,g\underline{x}^N) \in A$, we have with $v^{\alpha} = g\underline{x}^{\alpha}$, $\alpha = 1,\dots,N$,

$$P_{\theta}(g^{-1}A)$$

$$= \int_{g^{-1}A}(2\pi)^{-\frac{1}{2}Np}|\Sigma|^{-\frac{1}{2}N}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\sum_{\alpha=1}^{N}(\underline{x}^{\alpha}-\underline{\mu})(\underline{x}^{\alpha}-\underline{\mu})'\}\prod_{\alpha=1}^{N}d\underline{x}^{\alpha}$$

$$= \int_{A}(2\pi)^{-\frac{1}{2}Np}|g\Sigma g'|^{-\frac{1}{2}N}\exp\{-\tfrac{1}{2}\operatorname{tr}(g\Sigma g')^{-1}\sum_{\alpha=1}^{N}(\underline{y}^{\alpha}-g\underline{\mu})(\underline{y}^{\alpha}-g\underline{\mu})'\}\prod_{\alpha=1}^{N}d\underline{y}^{\alpha}$$

$$= P_{\bar{g}\theta}(A)\,.$$

The fact that $\bar{g}\,\Omega_{H_0} = \Omega_{H_0}$, $\bar{g}\,\Omega_{H_1} = \Omega_{H_1}$ is evident.

We are now concerned with the problem of characterizing the invariant tests. To do that we need the following definition.

Definition 2.2. (Invariant and maximal invariant functions): A function $T(x)$ defined on $\mathcal{X}$ is invariant under the group of transformations g operating from the left on $\mathcal{X}$ onto itself if $T(x) = T(gx)$ for all $x \in \mathcal{X}$ and for all $g \in G$. $T(x)$ is called a maximal invariant under G if it is invariant under G, and if $T(x) = T(y)$ for x,y belonging to $\mathcal{X}$ then there exists a $g \in G$ such that $y = gx$.

To investigate the behaviour of the maximal invariant with respect to the group G in the sample space, let for $x \in \mathcal{X}$

$$0(x) = \{gx: g \in G\}.$$

$0(x)$ is called the orbit of x with respect to G. A function $T(x)$ is a maximal invariant with respect to the group G if and only if it is constant on each orbit and it takes different values on different orbits. All maximal invariants are equivalent in the sense that their sets of constancy coincide.

Let $\phi(x)$, $x \in \mathcal{X}$ be a statistical test of H_0 against H_1, that is, $\phi(x)$ is the probability of rejecting H_0 when x is observed. For a non-randomized test $\phi(x)$ takes values 0 or 1. Suppose $\phi(x)$ is invariant under a group G of transformations g operating from the left on $\mathcal{X}$. The following theorem gives a useful characterization of $\phi(x)$ in terms of the maximal invariant $T(x)$.

Theorem 2.16. A statistical test $\phi(x)$ is invariant under G if and only if there exists a function h such that $\phi(x) = h(T(x))$.

Proof. Obviously if $\phi(x) = h(T(x))$ then for $x \in \mathcal{X}$, $g \in G$.

$$\phi(gx) = h(T(gx)) = h(T(x)) = \phi(x).$$

Conversely if $\phi(x)$ is invariant under G and $T(x) = T(y)$ for $x, y \in \mathcal{X}$, there exists a $g \in G$ such that $y = gx$ and therefore $\phi(x) = \phi(y)$.

Q.E.D.

In general h may not be a Borel measurable function. However if the range of T is Euclidean and T is Borel measurable then h is a Borel measureable function (Blackwell, 1956).

The induced group $\bar{G}$ operates from the left on Ω onto itself. We can thus, similarly, define a maximal invariant $v(\theta)$ on Ω with respect to the induced group $\bar{G}$. The following theorem asserts that the principle of invariance identifies all parameter points that are equivalent with respect to G.

Theorem 2.17. The distribution of the maximal invariant $T(X)$ with values in the space $\mathcal{Y}$ where X is a random variable with values in $\mathcal{X}$ depends on Ω only through $v(\theta)$.

Proof. Suppose that $v(\theta_1) = v(\theta_2)$ for θ_1, θ_2 in Ω. Since $v(\theta)$ is a maximal invariant on Ω under $\bar{G}$, there exists a $\bar{g} \varepsilon \bar{G}$ such that $\bar{\theta}_2 = \bar{g}\theta_1$. Let $\mathcal{B}$ be the σ-algebra in $\mathcal{Y}$. Then for $B \varepsilon \mathcal{B}$, by (2.31),

$$\begin{aligned} P_{\theta_1}\{T(X) \varepsilon B\} &= P_{\theta_1}\{T(gX) \varepsilon B\} \\ &= P_{\bar{g}\theta_1}\{T(X) \varepsilon B\} \\ &= P_{\theta_2}\{T(X) \varepsilon B\} \end{aligned}$$

Q.E.D.

Example 2.7. In Example 2.4 the maximal invariant in the sample space is $|t|$ where

$$T = \sqrt{N}\,\bar{X} \Big/ \sqrt{\tfrac{1}{n-1}\Sigma(x_i-\bar{x})^2}$$

and the corresponding maximal invariant in the parametric space is $|\lambda|$ where $\lambda = \sqrt{n}\,\mu/\sigma$, T has the Student-t distribution with $n-1$ degrees of freedom and its probability density function $f_T(t|\lambda)$ is given in (2.9). Under H_0, $\lambda = 0$ and under H_1, $|\lambda| > 0$. The probability density function of $|T|$ is given by, for $t > 0$

$$f_{|T|}(t|\lambda) = f_T(t|\lambda) + f_T(-t|\lambda).$$

Thus

$$\frac{f_{|T|}(t|\lambda)}{f_{|T|}(t|0)} = \exp\{-\tfrac{1}{2}\lambda^2\} \sum_{j=o}^{\infty} \frac{(\lambda^2)^{\frac{1}{2}j}\,\Gamma(\tfrac{1}{2}(n+j+1))}{j!\,\Gamma(\tfrac{1}{2}(n+1))}\,\frac{2^{\frac{1}{2}j}\,t^j}{(n+t^2)^{\frac{1}{2}j}} \qquad (2.42)$$

which is an increasing function of t for $t > 0$. Using Neyman and Pearson's fundamental lemma the uniformly most powerful invariant test of H_0 against the alternatives $H_1 : \mu \neq 0$ rejects H_0 whenever $|t| \geq C$, where the constant C depends on the size α of the test.

Note. The uniformly most powerful invariant test of size α is the uniformly most powerful test among all size α invariant tests.

Example 2.8. In Example 2.5 the maximal invariant in the sample space is $|t|$ where

$$t = \left(\frac{1}{m} + \frac{1}{n}\right)^{-\frac{1}{2}} (\bar{x} - \bar{y}) \Big/ \left(\frac{ms_1^2 + ns_2^2}{m+n-2}\right)^{\frac{1}{2}}$$

and the corresponding maximal invariant in the parametric space is $|\lambda|$ where

$$\lambda = \left(\frac{1}{m} + \frac{1}{n}\right)^{-\frac{1}{2}} (\mu_1 - \mu_2) / \sigma .$$

T has the Student-t distribution with $m + n-2$ degrees of freedom. Under H_0, $\lambda = 0$ and under H_1, $|\lambda| > 0$. Thus from the above example we conclude that for testing H_0 against H_1 the Student-t test which rejects H_0 whenever $|t| \geq C$, the constant C depending on the size α of the test, is uniformly most powerful invariant.

Example 2.9. Consider Example 2.6. Let

$$T(x) = N(N-1)\, \underline{\bar{x}}'\, s^{-1}\, \underline{\bar{x}} .$$

Since for $g \in G_\ell(p)$, the transformation $\underline{x}^\alpha \rightarrow g\underline{x}^\alpha$, $\alpha = 1, \ldots, N$, induces in the space of $(\underline{\bar{X}}, S)$ a transformation which transforms

$$(\underline{\bar{X}}, S) \rightarrow (g\underline{\bar{X}}, gSg'),$$

$$\begin{aligned}
T(gx) &= N(N-1)\, (g\underline{\bar{x}})'\, (gsg')^{-1}\, g\underline{\bar{x}} \\
&= N(N-1)\, \underline{\bar{x}}' g' (g')^{-1} s^{-1} g^{-1} g\underline{\bar{x}} \\
&= N(N-1)\, \underline{\bar{x}}'\, s^{-1}\, \underline{\bar{x}} \\
&= T(x) .
\end{aligned}$$

Thus $T(x)$ is invariant under $G_\ell(p)$ acting from the left on $\mathcal{X}$ into itself. To see that it is a maximal invariant let us suppose that for $x, y \in \mathcal{X}$

$$T(x) = N(N-1)\ \bar{\underline{x}}'\ s^{-1}\ \bar{\underline{x}},\quad T(y) = N(N-1)\ \bar{\underline{y}}'\ s_1^{-1}\ \bar{\underline{y}}\ ,$$

$$T(x) = T(y)\ ,$$

where $y = (\underline{y}^1, \ldots, \underline{y}^N)$,

$$\bar{\underline{y}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underline{y}^\alpha,\quad s_1 = \sum_{\alpha=1}^{N} (\underline{y}^\alpha - \bar{\underline{y}})(\underline{y}^\alpha - \bar{\underline{y}})'\ .$$

Since $N > p$, s and s_1 are both positive definite with probability one and hence there exist nonsingular matrices $\underline{B}$ and $\underline{B}_1$ such that $s = \underline{BB}'$, $s_1 = B_1B_1'$. Thus

$$(\underline{B}^{-1}\bar{\underline{x}})'\ (\underline{B}\bar{\underline{x}}) = (\underline{B}_1^{-1}\bar{\underline{y}})'\,(\underline{B}_1^{-1}\ \bar{\underline{y}})\ .$$

This implies that there exists an orthogonal matrix θ such that $\underline{B}^{-1}\bar{\underline{x}} = \underline{\theta}\ \underline{B}_1^{-1}\ \bar{\underline{y}}$ or equivalently $\bar{\underline{y}} = \underline{B}_1\ \underline{\theta}\ \underline{B}^{-1}\ \bar{\underline{x}}$.

Writing $g = B_1\theta B^{-1}$ we get $\bar{\underline{y}} = g\ \bar{\underline{x}}$ which implies that $\underline{y}^\alpha = g\underline{x}^\alpha$, $\alpha = 1, \ldots, N$. Obviously $g \in G_\ell(p)$. Thus $T(x)$ is a maximal invariant under $G_\ell(p)$.

Similarly a maximal invariant on Ω under the induced group $G_\ell(p)$ is

$$v(\theta) = N\underline{\mu}'\ \Sigma^{-1}\ \underline{\mu} = \delta^2\quad \text{(say)}.$$

$T(x)$ is distributed as Hotelling's T^2 and its probability density function is given in Theorem 2.14. Under H_0, $\delta^2 = 0$ and under H_1, $\delta^2 > 0$. The ratio of the probability density function of T^2 when $\delta^2 = \delta_1^2$ (fixed) to that when $\delta^2 = 0$ is given by

$$\frac{\Gamma(\tfrac{1}{2}p)}{\Gamma(\tfrac{1}{2}N)} \exp\{-\tfrac{1}{2}\delta_1^2\} \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j\,\Gamma(\tfrac{1}{2}N+j)}{j!\,\Gamma(\tfrac{1}{2}p+j)} \left(\frac{t^2}{N+1+t^2}\right)^j$$

where t^2 is a value of T^2. For fixed N and p and given δ_1^2, this is a strictly increasing function of $t^2/(N-1+t^2)$ and hence of t^2. By using Neyman and Pearson's lemma we conclude that the test which rejects H_0 whenever $t^2 \geq C$, the constant C depending on the size of the test, is uniformly most powerful invariant for testing H_0 against H_1.

In the search for the optimum invariant test it is then necessary to find a maximal invariant statistic and its distribution. Sometimes it is convenient to obtain a maximal invariant statistic through a number of steps each corresponding to a subgroup of the original group G. The question is under what conditions (if any) the maximal invariant under G and the maximal invariant obtained in steps corresponding to different subgroups of G are the same. For simplicity let us assume that G is generated by two subgroups H and K (it is the smallest group containing H and K). It is left to the reader to verify that G consists of the totality of products $\prod_{i=1}^{m} h_i k_i$, $m = 1, 2, \ldots$, where $h_i \in H$, $k_i \in K$, $i = 1, \ldots, m$.

Theorem 2.18. Let $y = f(x)$, $x \in \mathcal{X}$ be a maximal invariant on $\mathcal{X}$ with respect to the subgroup H operating on $\mathcal{X}$ from the left such that for any $k \in K$, $f(x_1) = f(x_2)$ with $x_1, x_2 \in \mathcal{X}$ implies that $f(kx_1) = f(kx_2)$. If $z = T(y)$ is a maximal invariant in the space of y with respect to the group K^* of transformation k^* defined by

$$k^* f(x) = f(kx),$$

then $T(f(x))$ is a maximal invariant with respect to G.

Proof. Let

$$g = k_m h_m \ldots k_1 h_1$$

where $h_i \in H$, $k_i \in K$, $i = 1, \ldots, m$. Then for $x \in \mathcal{X}$

$$\begin{aligned} T(f(gx)) &= T(f(k_m h_m \ldots k_1 h_1 x)) \\ &= T(k_m^* f(h_m k_{m-1} h_{m-1} \ldots k_1 h_1 x)) \\ &= T(f(k_{m-1} h_{m-1} \ldots k_1 h_1 x) \\ &\cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot \\ &= T(f(x)). \end{aligned}$$

Thus $T(f(x))$ is invariant under G.

Let $x_1, x_2 \in \mathcal{X}$ and let $T(f(x_1)) = T(f(x_2))$. Writing $y_1 = f(x_1)$ and $y_2 = f(x_2)$ we get $T(y_1) = T(y_2)$. Since $T(y)$ is a maximal invariant with respect to K^*. There exists $k^* \in K^*$ such that $y_2 = k^* y_1$. Thus $f(x_2) = k^* f(x_1) = f(kx_1)$. Since f is a maximal invariant with respect to H, (2.43) implies that there exists an $h \in H$ such that $x_2 = hkx_1$. Obviously hk is an element of G. Hence the theorem.

Q.E.D.

2.4.1. Almost Invariance and Invariance: To study the relative importance of different test criteria we need to compare their power functions. Thus it is of our interest to study the implication of invariance of power functions of tests rather than of the tests themselves. Since the power functions of invariant tests depend only on the maximal invariant in the parametric space, invariant tests have invariant power functions. The converse of this fact, namely, if the power function of a test ϕ is invariant under the induced group $\bar{G}$, that is,

$$E_\theta \phi(g^{-1}X) = E_{g\theta}\phi(X),$$

then the test ϕ is invariant under G, does not always hold good. To examine it further, let us first define the concepts of equivalence to an invariant test and the almost invariant test.

Definition 2.3. A test $\psi(x)$ is equivalent to an invariant test $\phi(x)$, $x \in X$ with respect to the group G of transformations if $\phi(x) = \psi(x)$ for all $x \in X-N$ where $P_\theta\{N\} = 0$ for all $\theta \in \Omega$.

Definition 2.4. A test $\phi(x)$ is said to be almost invariant with respect to the group G of transformations if for $g \in G$, $\phi(x) = \phi(gx)$ for all $x \in X-N_g$ where $P_\theta\{N_g\} = 0$ for all $\theta \in \Omega$.

Theorem 2.19. If a test ψ is equivalent to an invariant test ϕ then it is almost invariant. Conversely, if G is countable then any almost invariant test ψ is equivalent to an invariant test.

Proof. Take

$$N_g = N \cup (g^{-1}N).$$

Now $x \in X-N_g$ implies that $x \in X-N$ and $gx \in X-N$. So for $x \in X-N_g$, $\psi(x) = \phi(x) = \phi(gx) = \psi(gx)$. Since $P_\theta\{g^{-1}N\} = P_{\bar{g}\theta}\{N\} = 0$, $P_\theta\{N_g\} = 0$.

Conversely let the group G be countable. Write $N = \underset{g \in G}{\cup} N_g$. Since ψ is almost invariant, for $g \in G$, $\psi(x) = \psi(gx)$ for all $x \in X-N_g$ with $P_\theta\{N_g\} \rightarrow 0$. Thus $P_\theta\{N\} = 0$ and $\psi(x) = \psi(gx)$ for $x \in X-N$. Now define $\phi(x)$ such that

$$\phi(x) = \begin{cases} 1 & , \text{ if } x \in N \\ \psi(x) & , \text{ if } x \in \mathcal{X}-N. \end{cases}$$

Obviously, $\phi(x)$ is an invariant function and $\psi(x)$ is equivalent to an invariant test.

Q.E.D.

If the group G is uncountable the converse does not hold in general. We shall now investigate conditions under which the converse holds.

Let $\mathcal{X}$ be the sample space and let $\mathcal{A}$ be the σ-field of subsets of $\mathcal{X}$. Suppose that G is a group of transformations operating from the left on $\mathcal{X}$ onto itself with corresponding σ-field $\mathcal{B}$. Let for any $A \in \mathcal{A}$ the set of pairs (x,g) such that $gx \in A$ belongs to $\mathcal{A} \times \mathcal{B}$. Suppose that there exists a σ-finite measure μ (that is for $B_1, B_2, \ldots,$ in $\mathcal{B}$ with $\cup B_i = G$ and $\mu(B_i) < \infty$ for all i) on G such that $\mu(B) = 0$ implies that $\mu(Bg) = 0$ for all $g \in G$. Then any almost invariant function on $\mathcal{X}$ with respect to G is equivalent to an invariant function with respect to G. For a proof of this result the reader is referred to Lehmann (1959), p.225. The requirement is satisfied in particular when

$$\mu(Bg) = \mu(B), \quad B \in \mathcal{B},\ g \in G. \tag{2.43}$$

In other words μ is a σ-finite right invariant measure. Such right invariant measures exist for a large number of groups.

Example 2.10. Let $G = E^N$ (Euclidean N-space) with addition as the group operation. The Lebesgue measure μ in the space of E^N is a right invariant measure. Since G is abelian, the right invariant measure is also left invariant, that is, $\mu(gB) = \mu(B)$.

Example 2.11. Let G be the positive half of the real line with multiplication as the group operation. The right invariant measure is given by

$$\mu(B) = \int_B \frac{dg}{g} .$$

Example 2.12. Let G be the full linear group, that is, the multiplicative group of $N \times N$ nonsingular matrices $g = (g_{ij})$ and let $dg = \prod_{i,j} dg_{ij}$. Since the Jacobian of the transformation

$g \to gh$ for g, h in G is $|h|^N$, the right invariant measure on G is given by

$$\mu(B) = \int_B \frac{dg}{|g|^N} .$$

Example 2.13. Let G be the group of affine transformations of the real line R onto itself, that is, $g \in G$ has the form (a,b), $a \neq 0$ such that $(a,b)x = ax + b$. The group operation is defined for $g_1 = (a_1,b_1), g_2 = (a_2,b_2)$ in G, by $g_1g_2 = (a_1a_2, a_1b_2 + b_1)$. The Jacobian of the transformation $g_1 \to g_1g_2$ is a_2. The right invariant measure μ on G is

$$\mu(B) = \int_B \frac{da\,db}{a} .$$

The left invariant measure is not equal to the right invariant measure and is given by

$$\mu(B) = \int_B \frac{da\,db}{a^2} .$$

Example 2.14. Let G be the group of affine transformations for $\underline{x} \in V^N$, $g = (C,\underline{b})$, $C \in G_\ell(N)$, $\underline{b} \in V^N$, $g\underline{x} = C\underline{x} + \underline{b}$. The group operation in this case is defined by

$$g_1g_2 = (C_1C_2, C_1\underline{b}_2 + \underline{b}_1)$$

where $g_1 = (C_1,\underline{b}_1)$, $g_2 = (C_2,\underline{b}_2)$. The right invariant measure is given by

$$\mu(B) = \int_B \frac{dC\,db}{|C|^N} .$$

Theorem 2.20. Assume that the problem of testing $H_o: \theta \in \Omega_{H_o}$ against the alternatives $H_1: \theta \in \Omega_{H_1}$, where Ω_{H_o} and Ω_{H_1} are disjoint subsets of Ω, satisfying $\Omega_{H_o} \cup \Omega_{H_1} = \Omega$, remains invariant under the group G of transformation g operating from the left on x and that (2.43) holds. Then if ϕ_o is uniformly most powerful among the class of all invariant tests then it is also uniformly most powerful among the class of all almost invariant tests for this problem.

Proof. If ψ is almost invariant, by (2.44), it is equivalent to an invariant test ϕ. The tests ϕ and ψ have the same power function. Hence ϕ_o is uniformly at least as powerful as ψ.

Q.E.D.

It is now evident that any almost invariant test with respect to a group G of transformations has an invariant power function with respect to the induced group $\bar{G}$ on Ω. The converse of this does not always hold. However, the converse holds in cases where before the application of invariance, the problem can be reduced to one based on sufficient statistics in the sample space whose distributions constitute a boundedly complete family. Thus in such cases with the assumption (2.43) invariant tests can be compared with their power functions. We shall use the term sufficient statistic to mean the minimal sufficient statistic.

Let T be a sufficient statistic for $\{P_\theta, \theta \varepsilon \Omega\}$and let the family of distributions $\{P_\theta^T, \theta \varepsilon \Omega\}$ of T be boundedly complete, that is, for bounded function $f(T)$ of T if

$$E_\theta f(T) \equiv 0 \tag{2.44}$$

identically for $\theta \varepsilon \Omega$ then $f(T) = 0$ almost everywhere with respect to the probability measure P_θ^T. Let $\psi(T)$ be any invariant test with respect to the group G of transformations operating from the left on the space of the sufficient statistic T as

$$gT(x) = T(gx).$$

Then for $g \varepsilon G$

$$E_\theta \psi(T) = E_\theta \psi(gT) = E_{\bar{g}\theta} \psi(T)$$

so that $\psi(T)$ has an invariant power function. Conversely if

$$E_\theta \psi(T) = E_{\bar{g}\theta} \psi(T),$$

then for $g \varepsilon G$

$$E_\theta \psi(T) = E_\theta \psi(gT). \tag{2.45}$$

Hence

$$E_\theta(\psi(gT) - \psi(T)) \equiv 0$$

identically for all $\theta \varepsilon \Omega$. Since the distribution of T is boundedly complete, (2.46) implies that $\psi(T) = \psi(gT)$ almost everywhere with respect to the probability measure $P_\theta T$.

For any test $\phi(x)$ defined on the original sample space x

$$\psi(t) = E_\theta(\phi(X) \mid T = t)$$

is also a test with the same power function as that of $\phi(x)$. Thus we conclude that if there exists a uniformly most powerful almost invariant test among all tests based on the sufficient statistic T then that test is uniformly most powerful among all tests based on the original observations whose power functions depend on the maximal invariant in the parametric space.

Example 2.15. Consider Example 2.4. Let us first show that the sufficient statistic $(\bar{X}, S^2)$ for (μ, σ^2) is boundedly complete. From Example 2.1, $\bar{X}, S^2$ are independently distributed and nS^2/σ^2 is distributed as the central chisquare with $n-1$ degrees of freedom. Thus the joint probability density function of $\bar{X}, S^2$ is given by

$$f_{\bar{X},S^2}(\bar{x}, s^2) = k(\sigma^2)^{-\frac{1}{2}n} \exp\{-\frac{1}{2\sigma^2}(ns^2 + n(\bar{x}+\mu)^2)\}(ns^2)^{\frac{1}{2}(n-3)},$$

where

$k = \sqrt{n}/2\pi\ 2^{\frac{1}{2}(n-1)}\ \Gamma(\frac{1}{2}(n-1))$. For any bounded function $f(\bar{X}, S^2)$

$$E(f(\bar{X}, S^2)) = k \int f(\bar{x}, s^2)(\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}(ns^2 + n(\bar{x}-\mu)^2\}\, d\bar{x}\, ds^2$$

Write

$$\frac{1}{\sigma^2} = 1-2\theta \quad \text{and} \quad (1-2\theta)^{-1} t = \mu.$$

Then

$$E(f(\bar{X}, S^2)) = k \int f(\bar{x}, s^2)(1-2\theta)^{\frac{1}{2}n}(ns^2)^{\frac{1}{2}(n-3)}$$

$$\exp\{-\tfrac{1}{2}[(1-2\theta)(ns^2 + n\bar{x}^2) - 2nt\bar{x} + \frac{nt^2}{1-2\theta}]\}\, d\bar{x}\, ds^2 . \tag{2.46}$$

If $Ef(\bar{X}, S^2) = 0$ identically for all θ, then from (2.46) we get

$$k \int f(\bar{x}, ns^2 + n\bar{x}^2 - n\bar{x}^2)(ns^2)^{\frac{1}{2}(n-3)}$$

$$\exp\{-\tfrac{1}{2}(ns^2 + n\bar{x}^2) + \theta(ns^2 + n\bar{x}^2) + nt\bar{x}\}\, d\bar{x}\, ds^2$$

$$\equiv 0,$$

identically in θ and t. Treating this as the Laplace transform of

$$f(\bar{x}, ns^2+n\bar{x}^2-n\bar{x}^2)\ k(ns^2)^{\frac{1}{2}(n-3)}\ \exp\{-\tfrac{1}{2}(ns^2+n\bar{x}^2)\}$$

with respect to variables $n\bar{x}$, $ns^2+n\bar{x}^2$, we conclude that $f(\bar{x},s^2) = 0$ except for a set of values of $(\bar{X},S^2)$ with probability measure 0. So the distribution of $(\bar{X},S^2)$ is boundedly complete.

Suppose that for testing $H_0: \mu = 0$ against alternatives $H_1: \mu \neq 0$, the power of the test $\psi(\bar{x},s^2)$ depends only on $\lambda=\sqrt{n}\ \mu/\sigma$. Then for $g \in G$ (note $gx_i=ax_i$)

$$E_{\mu,\sigma^2}(\psi(\bar{X},S^2)) = E_{a\mu,a^2\sigma^2}(\psi(\bar{X},S^2))$$

$$= E_{\mu,\sigma^2}(\psi(a\bar{X},\ a^2S^2)).$$

So

$$E_{\mu,\sigma^2}[\psi(\bar{X},S^2) - \psi(a\bar{X},a^2S^2)] \equiv 0$$

identically in (μ,σ^2). This implies that

$$\psi(\bar{X},S^2) = \psi(a\bar{X},a^2S^2)$$

almost everywhere. Thus from Example 2.7 we conclude that of all tests of $\mu = 0$ with powers depending on λ, the Student t-test which reject H_0 whenever $|t| \geq C$ is uniformly most powerful invariant.

Example 2.16. Consider Example 2.6. Let us first establish that the statistic $(\underline{\bar{X}},S)$ is sufficient for $(\underline{\mu},\Sigma)$. From Theorem 2.12 the joint probability density function of $\underline{\bar{X}},S$ is given by

$$C|\Sigma|^{-\frac{1}{2}N}\ |s|^{\frac{1}{2}(N-p-2)}\ \exp\{-\tfrac{1}{2}\mathrm{tr}\ \Sigma^{-1}(s+N(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})')\},$$

where

$$C^{-1} = 2^{\frac{1}{2}Np}\ \pi^{p(p+1)/4}\ N^{\frac{1}{2}p}\ \prod_{i=1}^{p}\ \Gamma(\tfrac{1}{2}(N-i)).$$

For any real valued function $f(\underline{\bar{X}},S)$ of $(\underline{\bar{X}},S)$

$$E(f(\underline{\bar{X}},S)) = C\int f(\underline{\bar{x}},s)\,[|\Sigma|^{-\frac{1}{2}N}\ |s|^{\frac{1}{2}(N-p-2)}]$$

$$\times\ \exp\{-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1}s + N(\underline{\bar{x}}-\underline{\mu})'\ \Sigma^{-1}(\underline{\bar{x}}-\underline{\mu}))\}\ d\underline{\bar{x}}\ ds \qquad (2.47)$$

where $d\underline{\bar{x}} = \prod_{i=1}^{p} d\bar{x}_i$, $ds = \prod ds_{ij}$.

Write

$$\Sigma^{-1} = I - 2\theta \ , \quad \underline{\mu} = (I-2\theta)^{-1}\underline{\alpha} \ ,$$

where I is the $p \times p$ identity matrix and θ is a real symmetric matrix. If $E(f(\underline{\bar{X}},S)) \equiv 0$ identically for all $\underline{\mu}, \Sigma$, then from (2.47) we get

$$C \int f(\underline{\bar{x}},s) \ |I - 2\theta|^{\frac{1}{2}N} \ |s|^{\frac{1}{2}(N-p-2)}$$
$$\times \ \exp\{-\tfrac{1}{2}\mathrm{tr}((I-2\theta)(s+N\underline{\bar{x}}\,\underline{\bar{x}}') - 2N\underline{\alpha}' + N\underline{\alpha}'(I-2\theta)^{-1}\underline{\alpha})\} \ d\underline{\bar{x}} \ ds$$
$$= 0.$$

or equivalently,

$$C \int f(\underline{\bar{x}}, \ s+N\underline{\bar{x}}\,\underline{x}' - N\underline{\bar{x}}\,\underline{\bar{x}}') \ |s|^{\frac{1}{2}(N-p-2)}$$
$$\times \ \exp\{-\tfrac{1}{2}\mathrm{tr}(s+N\underline{\bar{x}}\,\underline{\bar{x}}') + \mathrm{tr}\theta\,(s+N\underline{\bar{x}}\,\underline{\bar{x}}') + N\underline{\alpha}'\underline{\bar{x}}\} \ d\underline{\bar{x}} \ ds$$
$$\equiv 0, \tag{2.48}$$

identically for all θ and $\underline{\alpha}$. Identifying (2.48) as the Laplace transform of

$$Cf(\underline{\bar{x}}, \ s+N\underline{\bar{x}}\,\underline{\bar{x}}' - N\underline{\bar{x}}\,\underline{\bar{x}}') \ |s|^{\frac{1}{2}(N-p-2)} \ \exp\{-\tfrac{1}{2}\mathrm{tr}(s+N\underline{\bar{x}}\,\underline{\bar{x}}')\}$$

with respect to variables $N\underline{\bar{x}}$, $s+N\underline{\bar{x}}\,\underline{\bar{x}}'$, we conclude from (2.48) that $f(\underline{\bar{x}},s) = 0$, except possibly for a set of values of $(\underline{\bar{X}},S)$ with probability measure 0. In other words, $(\underline{\bar{X}},S)$ is a boundedly complete sufficient statistic. Hence, if any test $\psi(\underline{\bar{X}},S)$ has invariant power function with respect to the induced group $\bar{G}_\ell(p)$, then

$$\psi(\underline{\bar{X}},s) = \psi(g\underline{\bar{x}}, \ gsg')$$

for $g \in G_\ell(p)$ almost everywhere, that is, ψ is almost invariant with respect $G_\ell(p)$. By Example 2.12 we conclude that of all tests of H_o: $\underline{\mu} = \underline{0}$ against alternatives H_1 : $\underline{\mu} \neq \underline{0}$ with power function depending on $N\,\underline{\mu}'\,\Sigma^{-1}\,\underline{\mu}$ the Hotelling's T^2-test is uniformly most powerful.

2.4.2. Sufficiency and Invariance: It is well known that some simplification is introduced in a testing problem by characterizing statistical tests as functions of sufficient statistics and thus reducing the dimension of the sample space to that of the space of the sufficient statistic without shrinking the parametric space. On the other hand invariance by reducing the dimension of the sample space to that of the space of the maximal invariant shrinks also the parametric space. Naturally we now ask the following question: Is it possible to use both sufficiency and invariance simultaneously and if so in what order? Under certain conditions the reduction of the sample space can be done by using both sufficiency and invariance and the order is immaterial in such cases. The reader is referred to Hall, Wijsman and Ghosh (1965) for these conditions and some related results. However, one can avoid the task of verifying these conditions by replacing the sample space by the space of the sufficient statistic before looking for the group of transformation which leaves the problem invariant and then look for the group of transformations on the space of the sufficient statistic that leaves the problem invariant.

2.4.3. Invariance and Unbiasedness: In testing statistical hypotheses the principle of unbiasedness plays an important role in deriving a suitable test statistic in complex situations involving composite hypotheses.

A size α test $\phi(x)$ is said to be unbiased for testing $H_o: \theta \in \Omega_{H_o}$ against alternatives $H_1: \theta \in \Omega_{H_1}$ if $E_\theta\, \phi(X) \geq \alpha$ for $\theta \in \Omega_{H_1}$.

In many complex testing problems the principle of unbiasedness and the principle of invariance seem to complement each other in the sense that each is successful in cases where the other is not. However, as we will demonstrate below, if both principles can be applied successfully they lead to the same (almost everywhere) optimum tests.

Consider the problem of testing $H_o: \theta \in \Omega_{H_o}$ against alternatives $H_1: \theta \in \Omega_{H_1}$. Let us assume that it is invariant under the group G of transformation g and let C_α be the class of all size α unbiased tests $(0 < \alpha < 1)$. For any test $\phi(x), x \in \mathcal{X}$ let us define the test ϕg by $\phi g(x) = \phi(gx)$, $g \in G$, $x \in \mathcal{X}$. Obviously $\phi \in C_\alpha$ if and only if $\phi g \in C_\alpha$. Hence if ϕ^* is the unique (upto a measure 0) uniformly most powerful unbiased test for this problem then

$$\begin{aligned}
E_\theta(\phi^* g(X)) &= E_\theta(\phi^*(gX)) \\
&= E_{\bar{g}\theta}(\phi^*(X)) \\
&= \sup_{\phi \in C_\alpha} E_{\bar{g}\theta}(\phi(X)) \\
&= \sup_{\phi g \in C_\alpha} E_\theta(\phi(gX)) \\
&= \sup_{\phi g \in C_\alpha} E_\theta(\phi g(X)) \\
&= E_\theta(\phi^*(X)).
\end{aligned}$$

Thus ϕ^* and $\phi^* g$ have the same power function. From the assumption of uniqueness ϕ^* is almost invariant. It may be remarked that if the sufficient statistic is complete, the uniformly most powerful unbiased test in terms of the statistic is unique. Therefore if there exists a uniformly most powerful almost invariant test ϕ^{**} we have

$$E_\theta \phi^{**}(X) \geq E_\theta \phi^*(X), \quad \theta \in \Omega_{H_1}. \tag{2.49}$$

Comparing ϕ^{**} with the trivial unbiased test $\phi(x) = \alpha$ for all $x \in \mathcal{X}$ we conclude that ϕ^{**} is unbiased and hence for $\theta \in \Omega_{H_1}$.

$$E_\theta \phi^{**}(X) \leq E_\theta \phi^*(X). \tag{2.50}$$

From (2.49) and (2.50) it follows that ϕ^* and ϕ^{**} have the same power function. From the uniqueness of ϕ^* we conclude that $\phi^* = \phi^{**}$ almost everywhere.

Example 2.17. Let $X_1, \ldots, X_N$ be a random sample of size N from a normal population with mean μ and variance σ^2. We are interested in finding the uniformly most powerful unbiased test of $H_o: \sigma^2 = \sigma_o^2$ against $H_1: \sigma^2 \neq \sigma_o^2$ where σ_o^2 is fixed. Write $s^2 = \sum_{i=1}^{N} (X_i - \bar{X})^2$.

The size α test which rejects H_o whenever $s^2/\sigma_o^2 \le C_1$ or $s^2/\sigma_o^2 \ge C_2$ where the constants C_1, C_2 are given by

$$\int_{C_1}^{C_2} f_{S^2}(s^2|\sigma_o^2)\,ds^2 = \frac{1}{N-1}\int_{C_1}^{C_2} s^2 f_{S^2}(s^2|\sigma_o^2)\,ds^2 = 1-\alpha$$

with

$$f_{S^2}(s^2|\sigma^2) = 2^{-\frac{1}{2}(N-1)}(\sigma^2)^{-\frac{1}{2}(N-1)}\exp\{-\tfrac{1}{2}\frac{s^2}{\sigma^2}\}(s^2)^{\frac{1}{2}(N-3)} \qquad (2.51)$$

is uniformly most powerful unbiased for testing H_o against H_1.

This problem remains invariant under the group G of translations, that is, $X_i \to X_i + a$, $i = 1,\ldots,N$. The maximal invariant in the sample space of $(X_1,\ldots,X_N)$ is S^2. From (2.51) it follows that no uniformly most powerful invariant test exists for this problem.

Example 2.18. Consider Example 2.7. The symmetric student t-test, which is uniformly most powerful invariant, is also uniformly most powerful unbiased for this problem.

Example 2.19. Let $(X_1,\ldots,X_m)$ and $(Y_1,\ldots,Y_n)$ be random samples from two normal populations with means μ_1,μ_2 and variances σ_1^2,σ_2^2 respectively and let

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i,\quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,\quad S_1^2 = \sum_{i=1}^{m}(X_i-\bar{X})^2,\quad S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2.$$

The problem of testing $H_o: \sigma_2^2/\sigma_1^2 \le k$ against alternatives $H_1: \sigma_2^2/\sigma_1^2 > k$, where k is fixed, remains invariant against the group of transformations

$$X_i \to aX_i + b_1,\ i = 1,\ldots,m,$$

$$Y_j = aY_j + b_2,\ j = 1,\ldots,n.$$

A maximal invariant in the sample space is

$$Z = \frac{m-1}{n-1}\,\frac{S_2^2}{S_1^2}\ .$$

From Theorem 2.2, with $\delta = \sigma_2^2/\sigma_1^2$,

$$f_Z(z|\delta) = \frac{n-1}{m-1}(\delta)^{\frac{1}{2}(m-3)}$$

$$\times \frac{\Gamma(\frac{1}{2}(m+n-2))}{\Gamma(\frac{1}{2}(m-1))\ \Gamma(\frac{1}{2}(n-1))} \quad \frac{(\frac{n-1}{m-1}z)^{\frac{1}{2}(n-3)}}{(\delta+\frac{n-1}{m-1}z)^{\frac{1}{2}(m+n-2)}}$$

for $z \geq 0$. Thus the distribution of Z depends only on δ and for $\delta < \delta'$ $f_Z(z|\delta)$, $f_Z(z|\delta')$ are distinct and $[f_Z(z|\delta')/f_Z(z|\delta)]$ is a nondecreasing function of z. So the test which rejects H_o whenever $z \geq C$, the constant C depending on the size α of the test, is uniformly most powerful invariant. It is also uniformly most powerful unbiased for testing H_o against H_1.

Example 2.20. Consider Example 2.9. For testing H_o against H_1, Hotelling's T^2 test is uniformly most powerful invariant. Whether it is also uniformly most powerful unbiased or not is still an open problem.

2.4.4. Invariance and Optimum Tests: Apart from the fact that the performance of an invariant test is independent of the nuisance parameters, powerful support for invariance comes from the famous unpublished work of Hunt and Stein towards the end of the Second World War which asserts that under certain conditions on the group G of transformations leaving the testing problem invariant there exists an invariant test which is minimax among all size α tests.

It is well known that given any test function on the sample space we can always replace it by a test depending only on the sufficient statistic such that both have the same power function. Such a result is too strong to expect from the maximal invariant statistic on the sample space. The appropriate weakening of the property and the condition under which it holds constitute the Hunt-Stein Theorem which states that if the problem of testing $H_o: \theta \in \Omega_{H_o}$ against $H_1: \theta \in \Omega_{H_1}$ remains invariant under the group G of transformations then given any test function ϕ on the sample space, there exists an invariant test ψ such that

$$\sup_{\theta\epsilon\Omega_{H_o}} E_\theta\psi \le \sup_{\theta\epsilon\Omega_{H_o}} E_\theta\phi$$

$$\inf_{\theta\epsilon\Omega_{H_1}} E_\theta\phi \le \inf_{\theta\epsilon\Omega_{H_1}} E_\theta\psi \qquad (2.52)$$

In other words ψ performs at least as well as ϕ in worst possible cases.

Theorem 2.21. (Hunt-Stein) Let $\mathcal{B}$ be a σ-field of subsets of G such that for any A ϵ $\mathcal{A}$ (a σ-field of subsets of the sample space $\mathcal{X}$) the set of pairs (x,g) with $gx \ \epsilon\ A$ is in $\mathcal{A} \times \mathcal{B}$ and for any B ϵ $\mathcal{B}$ and g ϵ G the set Bg is in $\mathcal{B}$. Suppose that there exists a sequence of probability distributions v_n over $(G,\mathcal{B})$ which is asymptotically right invariant, that is

$$\lim_{n\to\infty} |v_n(Bg) - v_n(B)| = 0. \qquad (2.53)$$

Then given any test ϕ there exists an almost invariant test ψ with respect to G satisfying (2.5).

We refer to Lehmann (1959) Chapter 8 for the proof and other discussions in this context. Giri, Kiefer and Stein(1963) and Kiefer(1966) are also referred to in this context. Thus, if for a given problem which is invariant with respect to a group G of transformation, Hunt-Stein condition is satisfied, then the uniformly most powerful invariant test will automatically be minimax, that is, will maximize the minimum power over Ω_{H_1}.

The group of translations of the real line, the group G in Example 2.4, the multiplicative group of orthogonal matrices and the multiplicative group of nonsingular lower triangular matrices are examples of some groups which satisfy the condition of the Hunt-Stein theerem. The full linear group $G_\ell(p)$ of nonsingular matrices does not satisfy the Hunt-Stein condition for $p \ge 2$.

Thus the symmetric Student t-test of Example 2.4 is minimax for testing $H_o: \mu = 0$ against the alternatives $H_1: \mu \ne 0$.

The Hunt-Stein theorem has been successfully used by Giri, Kiefer and Stein (1963), Giri and Kiefer(1964), Linnik, Pliss and Salaevskii(1966) and Salaevskii(1968) to prove the long time open problem of the minimax property of the Hotelling's T^2 test.

It may be remarked here that the conditions of the Hunt-Stein theorem, whether algebraic or topological, are almost entirely on the group and are non-statistical in nature. For verifying the admissibility of statistical tests through invariance the situation is more complicated. Aside from the trivial case of compact groups only the one-dimensional translation parameter case has been studied by Lehmann and Stein(1953). If G is finite or a compact group the most powerful invariant test is admissible. For the other groups the statistical structure plays an important role. The admissibility of Hotelling's T^2 test (and hence of the symmetric Student t-test) has been proved by Stein(1956) and Kiefer and Schwartz(1965) by using different methods.

2.4.5. Invariance and the Most Stringent Tests: Consider the problem of testing $H_o: \theta \in \Omega_{H_o}$ against the alternatives $H_1: \theta \in \Omega_{H_1}$ where Ω_{H_o} and Ω_{H_1} are disjoint subsets of Ω with $\Omega_{H_o} \cup \Omega_{H_1} = \Omega$. Let Q_α denote the class of all level α tests of H_o and let

$$\beta^*_\alpha(\theta) = \sup_{\phi \in Q_\alpha} E_\theta \phi, \ \theta \in \Omega_{H_1} .$$

$\beta^*_\alpha(\theta)$ is called the envelope power function and it is the maximum power that can be obtained at level α against the alternative θ

Definition 2.5. A test ϕ that minimizes

$$\sup_{\theta \in \Omega_{H_1}} (\beta^*_\alpha(\theta) - E_\theta(\phi))$$

is called the most stringent. In other words a most stringent test minimizes the maximum shortcomings.

If the testing problem is invariant under the group G of transformations satisfying the conditions of the Hunt-Stein theorem and there exists a uniformly most powerful almost invariant test ϕ^* with respect to G then ϕ^* is most stringent. As an example we cite the symmetric Student t-test, which is then the most stringent. The fact that Hotelling's T^2 test is most stringent has been established in Giri, Kiefer and Stein(1963).

For further details and supplementary reading in this context the reader is referred to Lehmann (1959, 1959a), Kiefer(1958), Giri and Kiefer(1964a).

Example 2.21. Let $X_1,\dots,X_N$ be independently normally distributed random variables with means $E(X_i)=\mu_i$, $i = 1,\dots,N$ and common variance σ^2 satisfying $\mu_{s+1} = \dots = \mu_N = 0$, $s < N$. We are interested in testing the hypothesis $H_o:\mu_1 = \dots = \mu_r = 0$, $r < s$. This problem remains invariant under three groups of transformations

$$G_1 : X_i \to X_i + a_i , \quad i = r+1,\dots,s$$

$$X_i \to X_i , \quad i = 1,\dots,r, s+1,\dots,N,$$

G_2 : orthogonal transformations of $X_1,\dots,X_r$,

G_3 : $X_i \to aX_i$, $a \neq 0$, $i = 1,\dots,N$.

With respect to G_1, G_2 and G_3 the uniformly most powerful invariant test rejects H_o when

$$\omega = \frac{\sum_{i=1}^{r} x_i^2/r}{\sum_{i=s}^{N} x_i^2/(N-s)} \geq C$$

where the constant C depends on the size α of the test. Under H_o, W has the central F-distribution with r and N-s degrees of freedom. The power of this test depends only on

$$\delta^2 = \sum_{i=1}^{r} \mu_i^2/\sigma^2 .$$

Since each of the groups G_1, G_2 and G_3 satisfies the condition of Hunt-Stein theorem this test is minimax and also most stringent.

Example 2.22. The following data were collected from an experiment on Jute Crop in Sagar-Bishnupur, West-Bengal, India. The weights in grams of green jute plants and dry jute fibre were recorded for twenty randomly selected individual plants for each of two different agricultural varieties grown under similar conditions. Let $\underline{x}^{\alpha} = (x_{1\alpha}, x_{2\alpha})'$ $\alpha = 1, \ldots, 20$, $\underline{y}^{\alpha} = (y_{1\alpha}, y_{2\alpha})'$, $\alpha = 1, \ldots, 20$, denote the sample values. The numerical values are given in Table 2.1.

Table 2.1

	Variety 1		Variety 2	
Plant No.	Weight in grams		Weight in grams	
	Green Plant	Dry Fibre	Green Plant	Dry Fibre
1	971	68	760	58
2	892	63	872	69
3	1125	70	1098	71
4	82	6	144	18
5	931	65	900	68
6	112	9	120	12
7	162	10	180	14
8	321	12	200	16
9	315	20	214	18
10	375	30	440	34
11	462	33	460	36
12	352	27	220	26
13	305	21	260	24
14	84	5	410	32
15	229	14	90	8
16	332	27	270	33
17	185	17	1080	70
18	703	53	888	63
19	872	62	209	18
20	740	65	700	49

Assume that these observations are samples from two bivariate normal populations with means $\underline{\mu} = (\mu_1, \mu_2)'$, $\underline{\nu} = (\nu_1, \nu_2)$ respectively and the same covariance matrix Σ.

Let us consider the problem of testing

(i) H_o: $\mu_2 = \nu_2$; (ii) $\underline{\mu} = \underline{\nu}$.

For Variety 1:

$$\underline{\bar{x}} = (\bar{x}_1, \bar{x}_2)' = (477.50, 33.85)$$

$$s_1 = (s^1_{ii}) = \sum_{\alpha=1}^{20} (\underline{x}^\alpha - \underline{\bar{x}})(\underline{x}^\alpha - \underline{\bar{x}})' = \begin{pmatrix} 2535681, & 149056.5 \\ 149056.5, & 10838.55 \end{pmatrix}$$

For Variety 2:

$$\underline{\bar{y}} = (\bar{y}_1 \bar{y}_2)' = (475.75, 36.85)',$$

$$s_2 = (s^2_{ij}) = \sum_{\alpha=1}^{20} (\underline{y}^\alpha - \underline{\bar{y}})(\underline{y}^\alpha - \underline{\bar{y}})' = \begin{pmatrix} 2239083, & 142218.2 \\ 142218.2, & 9350.55 \end{pmatrix}$$

$$s = s_1 + s_2 = \begin{pmatrix} 4374764, & 291274.7 \\ 291274.7, & 20189.10 \end{pmatrix}.$$

For (i). Student t statistic with 38 degrees of freedom is given by

$$t = \frac{\sqrt{380}(\bar{x}_2 - \bar{y}_2)}{\sqrt{s^1_{22} + s^2_{22}}} = -0.1415803 .$$

Hence we accept $H_o: \mu_2 = \nu_2$ at 5 percent level of significance.

For (ii). Hotelling's T^2 statistic with 38 degrees of freedom is

$$T^2 = 380(\underline{\bar{x}} - \underline{\bar{y}})' s^{-1} (\underline{\bar{x}} - \underline{\bar{y}}) = 0.0244102.$$

Hence at 5 percent level of significance we accept $H_o: \underline{\mu} = \underline{\nu}$.

EXERCISE 2

1. Show that if a quadratic form is distributed as a noncentral chisquare, the noncentrality parameter is the value of the quadratic form when the variables are replaced by their expected values

2. Show that if $f_X(x|\delta^2)$ is the probability density function of the noncentral chisquare or of a noncentral F random variable with the noncentrality parameter δ^2, then

$$\frac{f_X(x|\delta_1^2)}{f_X(x|\delta_o^2)}$$

is an increasing function of x for all $\sigma_o^2 < \sigma_1^2$.

3. Show that if X has central F distribution with r,N degrees of freedom then Y = rX/(rX+ N) has the central Beta distribution with ½r,½N degrees of freedom. Find also the distribution of log F.

4. Consider a chisquare (central) random variable X with N degrees ~~degrees~~ of freedom. Show that

$$\alpha = \sqrt{\frac{9}{2}},\ \beta = \frac{1}{3},\ \gamma = \frac{2}{9}$$

is a good choice for the rapid approach of the transformed variable

$$\alpha\sqrt{N}\ [(\frac{X}{N})^{\beta} + \frac{\gamma}{N} - 1]$$

to normality as $N \to \infty$.

5. Let $X_1,\ldots,X_N$ be independent and let each X_i be normally distributed with mean μ_i and variance σ_i^2. Show that if $X^* = (\sum_{i=1}^{N} \frac{1}{\sigma_i^2} X_i)/(\sum_{i=1}^{N} \frac{1}{\sigma_i^2})$, then $\sum_{i=1}^{N} (\frac{X_i - X^*}{\sigma_i})^2$ is distributed as $\chi^2_{N-1}(\delta^2)$ where $\delta^2 = 0$ when all μ_i's are equal.

6. Let S be distributed as Wishart $W_p(N-1,\Sigma)$. Show that for any non-null p-vector $\underline{L} = (\ell_1,\ldots,\ell_p)'$, $\underline{L}'S\underline{L}/\underline{L}'\Sigma\underline{L}$ is distributed as χ^2_{N-1}.

7. Giri (1965).

(a) Let $\underline{Z}$ be a p-variate complex normal random variable with mean $\underline{\alpha}$ and Hermitian positive definite covariance matrix Σ. Show that

$2\underline{Z}^*\ \Sigma^{-1}\underline{Z}$ is distributed as $\chi_p^2(2\underline{\alpha}^*\ \Sigma^{-1}\underline{\alpha})$.

(b) Let $\underline{Z}^{\alpha}$, $\alpha = 1,\ldots,N$ $(N > p)$ be a random sample of size N ~~fr~~ from a p-variate complex normal distribution with mean $\underline{\alpha}$ and Hermitian positive definite covariance matrix Σ, and let

$$\bar{\underline{Z}} = \frac{1}{N}\sum_{\alpha=1}^{N} \underline{Z}^{\alpha},\quad S = \sum_{\alpha=1}^{N} (\underline{Z}^{\alpha} - \bar{\underline{Z}})\ (\underline{Z}^{\alpha} - \bar{\underline{Z}})^*.$$

(i) Show that S, $\bar{\underline{X}}$ are independent.

(ii) The probability density function of S is

$$f_S(s) = K|\Sigma|^{-(N-1)}\ |s|^{N-p-1}\ \exp\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}s\}$$

where

$$K^{-1} = \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^{p} \Gamma(tN-i)).$$

(iii) Show that $T_c^2 = N(N-1)\underline{Z}^*S^{-1}\underline{Z}$ is distributed as $\chi^2_{2p}(2N\underline{\alpha}^*\Sigma^{-1}\underline{\alpha})/\chi^2_{2(N-1)}$, where $\chi^2_{2p}(.)$ and $\chi^2_{2(N-1)}$ are independent.

(iv) Among all size α tests of $H_o: \underline{\alpha} = \underline{0}$ against $H_1: \underline{\alpha} \neq \underline{0}$ with power functions depending on $N\underline{\alpha}^*\,\Sigma^{-1}\underline{\alpha}$ the likelihood ratio test which rejects H_o whenever

$$t_c^2 \geq k\ ,$$

where the constant k depends on the size α of the test is uniformly most powerful invariant.

8. Suppose that $\chi = E^N$, the Euclidean N-space .

(a) Let G be the group of translation of $(x_1,\ldots,x_N)'$ in χ to $(x_1+a,\ldots,x_N + a)'$. Show that

$$(x_1 - \bar{x},\ldots,x_N - \bar{x})'\ ,$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is a maximal invariant on χ under G.

(b) Let G be the group of orthogonal transformations of $\underline{x} \in \chi$. Show that $\underline{x}'\underline{x}$ is a maximal invariant under G.

9. Let $X_1,\ldots,X_N$ be a random sample form the normal population with mean μ and variance σ^2 and let G be the group scale of scale changes, that is, $g \in G$, $gX_i = aX_i$, $a > 0$. Show that the uniformly most powerful invariant test of $H_o\ \mu \leq 0$ against $H_1: \mu > 0$ when σ^2 is unknown rejects H_o whenever

$$\frac{\sqrt{N}\,\bar{x}}{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2} \geq C\ ,$$

where the constant C depends on the size α of the test. Show that it is also uniformly most powerful unbiased.

10. Let $X_1,\ldots,X_N$ be a random sample from the normal distribution with mean μ and variance σ^2. Show that among all tests of size α for $H_o: \sigma^2 \geq 1$ against $H_1: \sigma^2 < 1$, the size α test which rejects H_o whenever

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 < k ,$$

is uniformly most powerful invariant.

11. Let $X_1, \dots, X_m$, $Y_1, \dots, Y_n$ be random samples from two normal populations with means μ_1, μ_2 respectively and the common variance σ^2. Let

$$\bar{X} = \frac{1}{m} \sum_{i=1}^{m} X_i , \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i , \quad S^2 = \sum_{i=1}^{m} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$t = \frac{\left(\frac{1}{m} + \frac{1}{n}\right)^{-\frac{1}{2}} (\bar{y} - \bar{x})}{\sqrt{\frac{s^2}{m+n-2}}}$$

show that the uniformly most powerful invariant test of $H_0 : \mu_2 - \mu_1 \leq 0$ against $H_1 : \mu_2 - \mu_1 > 0$ rejects H_0 whenever $t \geq C$, the constant C depending on the size α of the test. Show that it is also uniformly most powerful unbiased.

12. Show that the distribution of Hotelling's T^2 statistics under alternative hypotheses remains unchanged if $\underline{\mu}$ is replaced by $(\delta, 0, \dots, 0)'$ and Σ by identity where $\delta^2 = N\underline{\mu}' \Sigma^{-1} \underline{\mu}$.

13. Consider Example 2.9.

(a) Show that $U = N\underline{\bar{X}}'(S + N\underline{\bar{X}}\underline{\bar{X}}')^{-1} \underline{\bar{X}} = N\underline{\bar{X}}'S^{-1}\underline{\bar{X}} \,/\, (T + N\underline{\bar{X}}'S^{-1}\underline{\bar{X}})$.

(b) Show that the uniformly most powerful invariant test of H_0 against H_1 rejects H_0 whenever $u \leq C$. Show that the probability density function of U under H_0 is given by

$$f_U(u) = \frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}p)\,\Gamma(\frac{1}{2}(N-p))} u^{\frac{1}{2}p-1} (1-u)^{\frac{1}{2}(N-p)-1}$$

$$\times \exp\{-\tfrac{1}{2}\delta^2\} \, \phi(\tfrac{1}{2}(N-p), \ \tfrac{1}{2}p; \ \tfrac{1}{2}u\,\delta^2)$$

where $\phi(a,b,x)$ is the confluent hypergeometric series given by

$$\phi(a,b,x) = 1 + \frac{a}{b}x + \frac{a(a+1)}{b(b+1)} \frac{x^2}{2!} + \dots .$$

REFERENCES

Blackwell, D. (1956). On a class of probability spaces, Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley.

Giri, N. (1965). On the complex analogues of T^2 - and R^2-tests. Annal. Math. Statist. 36, pp.664-670.

Giri, N. (1976). Multivariate statistical inference, Academic Press, N.Y.

Giri, N., Kiefer, J. and Stein, C. (1963). Minimax character of Hotelling's T^2 test in the simplest case. Annal. Math. Statist. 34, pp. 1524-25.

Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax tests. Annal. Math. Statist. 35, pp. 21-35.

Hall, W.J., Wijsman, R.A. and Ghosh, J.K. (1965). The relationship between sufficiency and invariance with application in sequential analysis. Annal.Math.Statist., 36, pp. 575-614.

Kiefer, J. (1966). Multivariate optimality results. Multivariate analysis, Edited by P.R. Krishnaiah, Academic Press, N.Y.

Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2 - and R^2-and other fully invariant tests for classical multivariate normal problems. Annal. Math. Statist. 36, pp. 747-770.

Lehmann, E.L. (1959). Testing statistical hypotheses. John Wiley and Sons, N.Y.

Lehmann, E.L. (1959a). Optimum invariant tests. Annal. Math. Statist. 30, pp. 881-884.

Lehmann, E.L. and Stein, C. (1953). The admissibility of certain invariant statistical tests involving translation parameter. Annal. Math. Statist., 24, pp. 473-479.

Linnik, Ju. V., Pliss, V.A. and Salaevskii (1966). Soviet Math. Dokl. 7, p. 719.

Salaevskii (1968). Minimax character of Hotelling's T^2-test. Soviet Math. Dokl., 9, pp. 733-735.

Stein, C. (1956). The admissibility of Hotelling's T^2-test. Annal. Math. Statist. 27, pp. 616-623.

Wilks, S.S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, pp. 471-494.

CHAPTER 3. UNIVARIATE GENERAL LINEAR HYPOTHESIS AND ANALYSIS OF VARIANCE

In this chapter we shall discuss two different forms of the univariate general linear hypothesis and their testing through analysis of variance techniques. The first one is the more abstract while the second one is more suitable for hypotheses concerning parameters in experimental design models, and will be called the regression model.

3.1. GENERAL LINEAR HYPOTHESIS

The first form of the general linear hypothesis concerns independent random variables $X_1,\dots,X_N$ having normal distributions with common variance σ^2 and means $E(X_i) = \mu_i$, $i = 1,\dots,N$. Let $\underline{\mu} = (\mu_1,\dots,\mu_N)'$, $\underline{X} = (X_1,\dots,X_N)'$. The mean vector $\underline{\mu}$ is assumed to lie in an s-dimensional subspace π_Ω of the N-dimensional Euclidean space E^N, $s < N$. Usually this assumption is stated in the form

$$\underline{\mu} = A\underline{\theta} \tag{3.1}$$

where A is a known $N \times s$ matrix of full rank s and $\underline{\theta}$ is an arbitrary s-dimensional vector. In this formulation π_Ω is the space spanned by the column vectors of A. The hypothesis H_o to be tested is that $\underline{\mu}$ lies in the s-r dimensional subspace π_ω of π_Ω with $r \leq s$. Usually the hypothesis is expressed as restricting $\underline{\theta}$ to satisfy

$$H_o : B\underline{\theta} = 0 ,$$

where B is an $r \times s$ matrix of full rank r. The general linear hypothesis can be put into a simpler form by transforming

$$\underline{X} \rightarrow \underline{Y} = (Y_1,\dots,Y_N)' = C\underline{X} \tag{3.2}$$

where $C = (c_{ij})$ is an $N \times N$ orthogonal matrix with the following properties.

The first s row vectors $\underline{c}_1,\dots,\underline{c}_s$ of C span π_Ω with $\underline{c}_{r+1},\dots,\underline{c}_s$ spanning π_ω. In other words the last N-s row vectors $\underline{c}_{s+1},\dots,\underline{c}_N$ of C are perpendicular to π_Ω and the first r rows

$\underline{c}_1,\dots,\underline{c}_r$ of C' are vectors in π_Ω which are perpendicular to π_ω. Clearly $\underline{Y}$ has the N-dimensional normal distribution with $E(\underline{Y}) = \underline{v} = (v_1,\dots,v_N)' = CA\theta$ and covariance matrix $\sigma^2 I$ and $\underline{v}$ satisfies the restriction $v_{s+1} = \dots = v_N = 0$. Under H_o: $v_1 = \dots = v_r = 0$. Thus by means of an orthogonal transformation of the above form we reduce the general linear hypothesis in the canonical form which concerns independent random variables $Y_1,\dots,Y_N$ having normal distributions with variance σ^2 and with means

$$E(Y_i) = \begin{cases} v_i \ , & i \le s \ , \\ 0 \ , & i > s \ , \end{cases} \tag{3.3}$$

and under H_o: $v_1 = \dots = v_r = 0$.

Example 3.1. Let us consider the geometrical interpretation of the above fact for the case N = 3 with

$$A = \begin{pmatrix} 1,0 \\ 1,0 \\ 0,1 \end{pmatrix} , \qquad \underline{\theta} = (\theta_1,\theta_2)' \ .$$

The rank of A is 2 and $\mu_1 = \theta_1$, $\mu_2 = \theta_1$, $\mu_3 = \theta_2$. So π_Ω is a subspace of E^3 where $\mu_1 = \mu_2$. Under H_o, $\theta_2 = \theta_1$ and π_ω is a subspace of π_Ω with $\mu_1 = \mu_2 = \mu_3$. Clearly the second row of C is $(1/\sqrt{3},\ 1/\sqrt{3},\ 1/\sqrt{3})$. The first row of C is a normalized vector which is a linear combination of (1,1,0) and (0,0,1) and is perpendicular to the second row. Hence it is given by $(1/\sqrt{6},\ 1/\sqrt{6},\ -2/\sqrt{6})$. Similarly the third row of C, which is perpendicular to the first and the second row is $(1/\sqrt{2},\ -1/\sqrt{2}, 0)$. The mean vector $\underline{v}$ is

$$\underline{v} = \left(\frac{2}{\sqrt{6}}(\theta_1 - \theta_2),\ \frac{2\theta_1 + \theta_2}{\sqrt{3}},\ 0 \right)' \ .$$

Geometrically we represent them in Figure 3.1.

Example 3.2. Let X_{ij}, $j = 1,\dots,N_i$, $i = 1,\dots,I$ be independent normally distributed random variables with common variance σ^2 and with means $E(X_{ij}) = \mu_i$, $j = 1,\dots,N_i$, $i = 1,\dots,I$. Here $\underline{X} = (X_{11},\dots,X_{1N_1},\dots,X_{I1},\dots,X_{IN_I})'$, $s = I$, $N = \sum_{i=1}^{I} N_i$. For the null hypothesis H_o: $\mu_1 = \dots = \mu_I$ $s-r = 1$. The matrix $A = (a_{ij})$ is given by

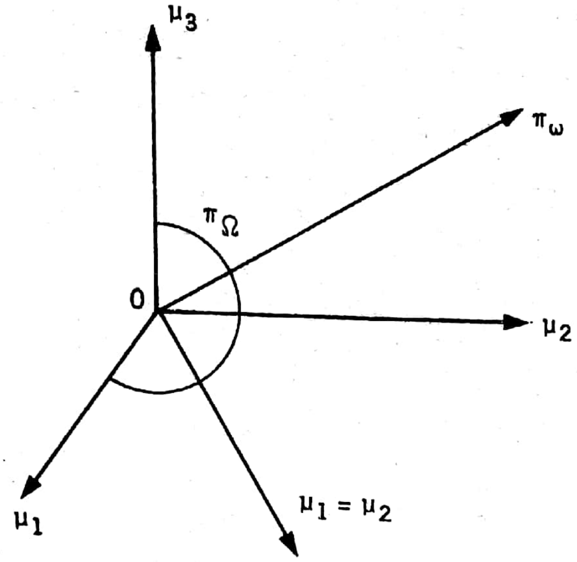

Figure 3.1

$$a_{ij} = \begin{cases} 1, & \text{if } N_1 + \dots + N_{j-1} < i \leq N_1 + \dots + N_j , \\ 0, & \text{otherwise.} \end{cases}$$

is spanned by the column vectors of A and π_ω is spanned by the single vector whose components are all equal to unity.

This is called the one-way classification. The following numerical data were collected from an experiment with four different varieties of wheat at the Indian Agricultural Research Institute, New Delhi. We shall use the data for further discussion with one-way classification.

3.1.1. Invariance of General Linear Hypothesis: The canonical form of the general linear hypothesis in terms of $Y_1, \dots, Y_N$ remains invariant under the group generated by the following 3 subgroups:

Group of translations G_1, translating

$\underline{Y} \to g\underline{Y} = (Y_1^*, \dots, Y_N^*)'$

with $Y_i^* = Y_i + a_i$, $i = r+1, \dots, s$, $Y_i^* = Y_i$ otherwise.

Group G_2 of orthogonal transformations of $Y_1, \dots, Y_r$.

Group G_3 of scale changes, changing $\underline{Y}$ to $g\underline{Y} = a\underline{Y}$ with $a \neq 0$.

Table 3.1. Weight in grams per 10 ears for fourteen randomly selected plants for each of four different varieties of wheat grown under similar conditions.

Plant No.	Varieties 1	2	3	4
1	13.7	18.0	16.5	16.2
2	13.9	17.6	16.1	16.6
3	17.7	16.3	14.6	19.3
4	17.3	14.2	11.6	17.2
5	13.8	12.7	12.8	11.8
6	15.7	12.1	8.9	11.6
7	16.3	15.3	17.4	12.9
8	13.1	14.8	12.6	17.0
9	13.6	15.0	15.9	17.2
10	13.6	14.0	12.1	15.9
11	17.6	9.6	14.0	18.4
12	16.7	17.7	13.4	18.7
13	18.2	17.3	15.1	13.8
14	11.8	17.6	17.9	14.9

The joint probability density function of $Y_1,\dots,Y_N$ is

$$(2\pi)^{-\frac{1}{2}N}(\sigma^2)^{-\frac{1}{2}N}\exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{s}(y_1-v_i)^2+\sum_{i=s+1}^{N}y_i^2\right\}\right] \tag{3.4}$$

Thus $(Y_1,\dots,Y_s, \sum_{i=s+1}^{N} Y_1^2)$ is a sufficient statistic for $(v_1,\dots,v_s,\sigma^2)$. So sufficiency reduces the problem to the consideration of $(Y_1,\dots,Y_s, \sum_{i=s+1}^{N} Y_i^2)$. In the space of this sufficient statistic a maximal invariant under G_1 is $(Y_1,\dots,Y_r, \sum_{i=s+1}^{N} Y_i^2)$, a maximal invariant under G_1, G_2 is $(U = \sum_{i=1}^{r} Y_i^2,\ V = \sum_{i=s+1}^{N} Y_i^2)$ and a maximal invariant under G_1, G_2 and G_3 is U/V. Thus the principle of invariance reduces the data to the single statistic U/V or equivalently to the statistic

$$W = \frac{N-s}{r} \frac{U}{V} .$$

Each of the 3 groups G_i, $i = 1, 2, 3$, of transformations induces in the parametric space the following group of transformations $\bar{G}_i$. $\bar{G}_i$ is the group of translations $\underline{v} \to g\underline{v} = \underline{v}^* = (v_1^*, \ldots, v_N^*)'$ with $v_i^* = v_i + a_i$, $i = r+1, \ldots, s$; $v_i^* = v_i$, $i = 1, \ldots, r$ and $v_i^* = 0$ otherwise, which leaves $(v_1^*, \ldots, v_r^*, \sigma^2)$ as the maximal invariant in the parametric space.

$\bar{G}_2$ is the group of orthogonal transformations of $(v_1, \ldots, v_r)$ which leaves $(\sum_{i=1}^{r} v_i^2 / \sigma^2)$ as the maximal invariant in the parametric space under $\bar{G}_1$ and $\bar{G}_2$.

$\bar{G}_3$ is the group of scale change $\underline{v} \to a\underline{v}$, $a \neq 0$, which leaves $\delta^2 = \sum_{i=1}^{r} v_i^2/\sigma^2$ as the maximal invariant in the parametric space under $\bar{G}_1, \bar{G}_2, \bar{G}_3$. The distribution of W depends only on δ^2 and has the noncentral F distribution with r, N-s degrees of freedom. From Example 2.21 the uniformly most powerful invariant test rejects $H_o: v_1 = \cdots = v_r = 0$ whenever

$$\omega \geq C , \tag{3.5}$$

where the constant C depends on the size α of the test. It is also uniformly most powerful among all tests with power functions depending only on δ^2. If $r = 1$ the test given in (3.5) reduces to the symmetric Student's-t test and is uniformly most powerful unbiased for testing H_o against $H_1: v_1 \neq 0$. On the other hand no uniformly most powerful unbiased test exists if $r > 1$.

3.1.2. Least Squares and the Canonical Form: In actual practice it is usually not very convenient to carry out explicitly the reduction of the general linear hypothesis to the canonical form. However, we can express the maximal invariant statistic W in terms of original observable random variables $X_1, \ldots, X_N$ by the theory of least squares. Let

$$\begin{aligned} S^2 &= (\underline{Y} - E(\underline{Y}))' (\underline{Y} - E(\underline{Y})) \\ &= \sum_{i=1}^{s} (Y_i - v_i)^2 + \sum_{i=s+1}^{N} Y_i^2 . \end{aligned} \tag{3.6}$$

Since $\sum_{i=s+1}^{N} y_i^2$ is the minimum value of S^2 without any restrictions on the values of $v_1,\ldots,v_s$ and the orthogonal transformation $\underline{y} = c\underline{x}$ leaves the distance unchanged, that is,

$$S^2 = (\underline{y} - E(\underline{Y}))'\,(\underline{y} - E(\underline{Y}))$$

$$= (\underline{x} - \underline{\mu})'\,(\underline{x} - \underline{\mu})$$

$$= \sum_{i=1}^{N} (x_i - \mu_i)^2 .$$

The minimum values of $\sum_{i=1}^{s} (y_i - v_i)^2 + \sum_{i=s+1}^{N} y_i^2$, under variations of $v_1,\ldots,v_s$ is also the minimum value of $(\underline{x} - \underline{\mu})'(\underline{x} - \underline{\mu})$ under variations of $\underline{\mu}$ in π_Ω. Thus if $\hat{\underline{\mu}} = (\hat{\mu}_1,\ldots,\hat{\mu}_N)'$ is the least squares estimate of $\underline{\mu}$ under π_Ω, then,

$$\sum_{i=s+1}^{N} y_i^2 = (\underline{x} - \hat{\underline{\mu}})'(\underline{x} - \hat{\underline{\mu}})$$

$$= \min_{\pi_\Omega} (\underline{x} - \underline{\mu})'(\underline{x} - \underline{\mu}) .$$

Similarly under π_ω the minimum value $\sum_{i=1}^{r} y_i^2 + \sum_{i=s+1}^{N} y_i^2$ of

$$\sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{s} (y_i - v_i)^2 + \sum_{i=s+1}^{N} y_i^2$$

can be expressed as

$$\sum_{i=1}^{r} y_i^2 + \sum_{i=s+1}^{N} y_i^2 = (\underline{x} - \hat{\hat{\underline{\mu}}})'(\underline{x} - \hat{\hat{\underline{\mu}}}) ,$$

where $\hat{\hat{\underline{\mu}}}$ is the least squares estimate of $\underline{\mu}$ under π_ω, that is,

$$(\underline{x} - \hat{\hat{\underline{\mu}}})'(\underline{x} - \hat{\hat{\underline{\mu}}}) = \min_{\pi_\omega} (\underline{x} - \underline{\mu})'(\underline{x} - \underline{\mu}) .$$

So

$$\omega = \frac{N-s}{r} \frac{(\hat{\underline{\mu}} - \hat{\hat{\underline{\mu}}})'(\hat{\underline{\mu}} - \hat{\hat{\underline{\mu}}})}{(\underline{x} - \hat{\underline{\mu}})'(\underline{x} - \hat{\underline{\mu}})} . \qquad (3.7)$$

Geometrically we represent them as in Figure 3.2. The vectors $\hat{\underline{\mu}}, \hat{\hat{\underline{\mu}}}$ are the projections of $\underline{x}$ on π_Ω, π_ω respectively.

Hence the vectors $\underline{x} - \hat{\underline{\mu}}$, $\underline{x} - \hat{\hat{\underline{\mu}}}$ are orthogonal.

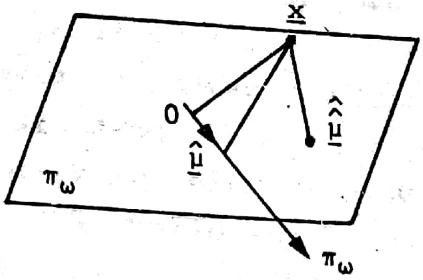

Figure 3.2

3.2. REGRESSION MODEL OF THE GENERAL LINEAR HYPOTHESIS

We shall now discuss a different formulation of the general linear hypothesis which is very appropriate for analysis of data from experimental designs. Let $X_1, \ldots, X_N$ be a sequence of N independent random variables with

$$E(X_i) = \sum_{j=1}^{p} \beta_j x_{ij},$$

$$\operatorname{Var}(X_i) = \sigma^2, \tag{3.8}$$

$i = 1, \ldots, N$, where $\beta_1, \ldots, \beta_p$ are unknown quantities and x_{ij}, $i=1, \ldots, N$, $j = 1, \ldots, p$ are known constant. The set of equations (3.8) is called a linear model for the expectations with independent covariance structure. In the vector notation $\underline{X} = (X_1, \ldots, X_N)'$ $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$ and the matrix $X = (x_{ij})$ of dimensions $N \times p$, we can rewrite (3.8) as

$$E(\underline{X}) = X\underline{\beta}$$

$$E(\underline{X} - X\underline{\beta})(\underline{X} - X\underline{\beta})' = \sigma^2 I. \tag{3.9}$$

We shall assume throughout this chapter that $N \geq p$. If the rank of X is p then the model is called a model of full rank. Linear models where the matrix X is not of full rank is of common occurrence in experimental designs. In this connection the matrix X is called the design matrix.

In experimental designs the unknown quantities $\beta_1, \ldots, \beta_p$ are called effects and the known constants x_{ij} are called countervariables which count the number of times the effect β_j occurs in the ith observation x_i. In most situations of experimental designs x_{ij} usually take the values 1 or 0 indicating the presence or the absence of the effect β_j in x_i. It is not always that x_{ij} take the values 1 or 0. The cases where x_{ij} are continuous variables are referred to as models of regression analysis. The mixed cases, where some x_{ij} are countervariables and the rest are continuous variables, are called models of analysis of covariance.

The unknown effects β_j may be constants or random variables. If all β_j are constants we call them parameters of the linear model and the model is called a fixed effect model. If for some j, $x_{ij} = 1$ for $i = 1, \ldots, N$, then the corresponding β_j is called an additive constant. If a model is such that all β_j are observable random variables, it is called a random-effect model. In a mixed effect model some β_j are parameters and some are random variables. We shall consider mainly the problem of making statistical inferences about the parameters in a fixed effect model.

3.2.1. Least Squares Estimate of $\underline{\beta}$: Let $x_1, \ldots, x_N$ be a sample of observations on $X_1, \ldots, X_N$. The least squares estimate of the parameters $\underline{\beta}$ in the linear model is the value $\underline{\bar{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)'$ of $\underline{\beta}$ for which

$$S^2 = (\underline{x} - X\underline{\beta})' (\underline{x} - X\underline{\beta}),$$

where $\underline{x} = (x_1, \ldots, x_N)'$, is minimum. Obviously given X, $\bar{\beta}_i$ is a function of $\underline{x}$ only. Differentiating S^2 with respect to β_k, the value $\underline{\bar{\beta}}$ of $\underline{\beta}$ which minimizes S^2 is given by

$$\sum_{i=1}^{N} x_{ik}(x_i - \sum_{j=1}^{p} \bar{\beta}_j x_{ij}) = 0, \; k = 1, \ldots, p . \tag{3.10}$$

In matrix notation we rewrite (3.10) as

$$X'X\underline{\bar{\beta}} = X'\underline{x} . \tag{3.11}$$

The equations

$$X'X\underline{\bar{\beta}} = X'\underline{x} \tag{3.12}$$

are called normal equations for the linear model and it is a nonhomogeneous system of linear equations in $\underline{\beta}$. If X has the full rank p then X'X has a unique inverse $(X'X)^{-1}$ and the unique least squares estimate of $\underline{\beta}$ is given by

$$\underline{\bar{\beta}} = (X'X)^{-1}X'\underline{x} \; . \tag{3.13}$$

In the general case the normal equations (3.11) always admit of a nontrivial solution, though, obviously it is not unique. This is because the vector $X'\underline{x}$ belongs to the column space of X' which is the same as the column space of X'X. If $(X'X)^{-}$ is any generalized inverse of (X'X), then a solution of (3.11) is given by (see Chapter 1)

$$\underline{\bar{\beta}} = (X'X)^{-} X' \underline{x} + (I-(X'X)(X'X)^{-})\underline{z} \; , \tag{3.14}$$

where $\underline{z}$ is any arbitrary p-vector.

Lemma 3.1. s^2 has a unique minimum for all solutions $\underline{\bar{\beta}}$ of (3.11).

Proof. Let $\underline{\bar{\beta}}$ be any solution of (3.11). Then

$$\begin{aligned} s^2 &= (\underline{x} - X\underline{\beta})'(\underline{x} - X\underline{\beta}) \\ &= (\underline{x} - X\underline{\bar{\beta}} + X(\underline{\bar{\beta}} - \underline{\beta}))'(\underline{x} - X\underline{\bar{\beta}} + X(\underline{\bar{\beta}} - \underline{\beta})) \\ &= (\underline{x} - X\underline{\bar{\beta}})'(\underline{x} - X\underline{\bar{\beta}}) + (\underline{\bar{\beta}} - \underline{\beta})'XX'(\underline{\bar{\beta}} - \underline{\beta}) \; . \end{aligned} \tag{3.15}$$

Since the second term in the above expression is nonnegative, s^2 attains its minimum value $(\underline{x} - X\underline{\bar{\beta}})'(\underline{x} - X\underline{\bar{\beta}})$ when $\underline{\beta} = \underline{\bar{\beta}}$ and this minimum value is unique for all solutions $\underline{\bar{\beta}}$ of (3.11).

A least square estimate of any linear parametric function $\underline{L}'\underline{\beta}$, where $\underline{L}=(\ell_1,\ldots,\ell_p)'$ is a nonnull p-vector, is defined by $\underline{L}'\underline{\bar{\beta}}$ where $\underline{\bar{\beta}}$ is a solution of (3.11). Obviously $\underline{L}'\underline{\bar{\beta}}$ is unique if X has full rank. We shall now investigate conditions under which a linear parametric function $\underline{L}'\underline{\beta}$ admits of a unique least squares estimate in the general case. To do this we need the following concept of estimable functions. In what follows we shall not distinguish between an estimate and the corresponding estimator. An estimator corresponding to any estimate is obtained by replacing $\underline{x}$ by $\underline{X} = (X_1,\ldots,X_N)'$ in the expression for the estimate.

Definition 3.1. A parametric function (or a parameter) is estimable if there exists an unbiased estimator of the parametric function (or the parameter) which is linear in the components of $\underline{X}$.

Such estimable functions are usually called linearly estimable. For convenience we prefer to call them simply estimable functions. The concept of estimable functions is due to Bose (1944). It is easy to observe that not all parametric functions are estimable. A condition of estimability of the parametric function is given in the following theorem.

Theorem 3.2. The linear parametric function $\underline{L}'\underline{\beta}$ is estimable if and only if the equations $X'X\underline{\lambda} = \underline{L}$ in $\underline{\lambda}$ has a solution.

Proof. If $\underline{L}'\underline{\beta}$ is estimable then there exists a linear combination $\underline{b}'\underline{X}$, where $\underline{b} = (b_1, \ldots, b_N)'$ is a real N-vector, such that,

$$E(\underline{b}'\underline{X}) = \underline{b}'X\underline{\beta} = \underline{L}'\underline{\beta} \quad \text{for all } \underline{\beta} .$$

Hence $X'\underline{b} = \underline{L}$. This implies that the rank of the vector subspace spanned by the column vectors of X equals the rank of the augmented matrix $(X'|\underline{L})$. In other words the rank of $X'X$ equals the rank of the augmented matrix $(X'X|\underline{L})$. Hence $X'X\underline{\lambda} = \underline{L}$ in $\underline{\lambda}$ has a solution.

Conversely suppose that $X'X\,\underline{\lambda} = \underline{L}$ has a solution for $\underline{\lambda}$. Writing $\underline{b} = X\underline{\lambda}$ we get $\underline{b}'\underline{X} = \underline{L}'$. So $\underline{b}'X\underline{\beta} = \underline{L}'\underline{\beta}$ for all $\underline{\beta}$. This implies that $\underline{b}'E(\underline{X}) = \underline{L}'\underline{\beta}$ and hence $\underline{L}'\underline{\beta}$ is estimable and $\underline{b}'\underline{X}$ is its linear unbiased estimator.

Q.E.D.

Comment. Since $X'X\underline{\lambda} = \underline{L}$ implies that $X'\underline{b} = \underline{L}$, the above theorem can be stated in terms of the condition that $\underline{L}$ is a linear combination of the column vectors of X'.

We now state the conditions under which a linear parametric function $\underline{L}'\underline{\beta}$ admits of a unique least squares estimate.

Theorem 3.3. A linear parametric function $\underline{L}'\underline{\beta}$ admits of a unique least squares estimate if and only if $\underline{L}'\underline{\beta}$ is estimable.

Proof. If $\underline{L}'\underline{\beta}$ is estimable, by Theorem 3.2, $\underline{L}$ belongs to the column space of X'X. If $\bar{\underline{\beta}}^{(1)}$ and $\bar{\underline{\beta}}^{(2)}$ are any two solutions of (3.11), then $(\bar{\underline{\beta}}^{(1)} - \bar{\underline{\beta}}^{(2)})$ is orthogonal to the column space of X'X. Hence $(\bar{\underline{\beta}}^{(1)} - \bar{\underline{\beta}}^{(2)})$ is orthogonal to $\underline{L}$ which implies that $L'\bar{\underline{\beta}}^{(1)} = \underline{L}'\bar{\underline{\beta}}^{(2)}$.

Conversely if $\underline{L}'\underline{\beta}$ admits of a unique least squares estimate it follows that any vector which is orthogonal to the column space of X'X is also orthogonal to $\underline{L}$. This means that $\underline{L}$ belongs to the column space of X'X and hence $\underline{L}'\underline{\beta}$ is estimable.

Q.E.D.

Another equivalent condition of estimability of the linear parametric function, which is easier to check, involves a generalized inverse of the matrix X'X and is given in the following theorem.

Theorem 3.4. The linear function $\underline{L}'\beta$ is estimable if and only if $L'(I-(X'X)^{-}(X'X)) = \underline{0}$ where $(X'X)^{-}$ is a generalized inverse of X'X.

Proof. By Theorem 3.3 $\underline{L}'\beta$ is estimable if and only if $\underline{L}'\beta$ admits of a unique least squares estimate. Since a solution of (3.11) is given by (3.14) and $\underline{z}$ is arbitrary, it follows that we must have

$$\underline{L}'(I - (X'X)^{-}(X'X)) = \underline{0}\,. \tag{3.16}$$

Q.E.D.

We shall now investigate the optimum properties of the least squares estimator of any linear parametric function. Since the joint distribution of $X_1, \ldots, X_N$ is not specified, in general it will not be possible to examine any optimum properties of the least squares estimator in relation to all possible estimators. Instead, we shall investigate its optimum property among all linear estimators, that is, estimators which are linear functions of $X_1, \ldots, X_N$.

Theorem 3.5. (Gauss-Markoff). If the linear parametric function $\underline{L}'\beta$ is estimable then the linear estimator $\underline{L}'\bar{\underline{\beta}}$ where $\bar{\underline{\beta}}$ is a solution of

$$X'X\underline{\beta} = X'\underline{X} \tag{3.17}$$

is best linear unbiased, that is, best (minimum variance) among the class of all unbiased linear estimators of $\underline{L}'\underline{\beta}$.

Note: A best linear unbiased estimator is often called BLUE.

Proof. Since $\underline{L}'\underline{\beta}$ is estimable, from Theorem 3.2 we get

$X'X\underline{\lambda} = \underline{L}$ for some $\underline{\lambda}$. Hence

$$\underline{L}'\bar{\underline{\beta}} = \underline{\lambda}'X'X\bar{\underline{\beta}} = \underline{\lambda}'\ X'\underline{X} \ .$$

Therefore,

$$\begin{aligned} E(\underline{L}'\bar{\underline{\beta}}) &= E(\underline{\lambda}'\ X'\underline{X}) \\ &= \underline{\lambda}'\ X'X\underline{\beta} \\ &= \underline{L}'\underline{\beta} \ . \end{aligned}$$

Let $\underline{b}'\underline{X}$ be any other linear unbiased estimator of $\underline{L}'\underline{\beta}$. Since $\underline{b}'\underline{X}$ is unbiased for $\underline{L}'\underline{\beta}$, we get $\underline{b}'X = \underline{L}'$. Now

$$\begin{aligned} \mathrm{Var}(\underline{b}'\underline{X}) &= \mathrm{Var}(\underline{b}'\underline{X} - \underline{L}'\bar{\underline{\beta}} + \underline{L}'\bar{\underline{\beta}}\ ) \\ &= \mathrm{Var}(\underline{b}'\underline{X} - \underline{\lambda}'X'\underline{X} + \underline{L}'\bar{\underline{\beta}}\ ) \\ &= \mathrm{Var}(\underline{b}'\underline{X} - \underline{\lambda}'X'\underline{X}) + \mathrm{Var}(\underline{L}'\bar{\underline{\beta}}) \\ &\geq \mathrm{Var}(\underline{L}'\bar{\underline{\beta}})\ , \end{aligned}$$

as

$$\begin{aligned} \mathrm{Cov}(\underline{b}'X - \underline{\lambda}'X'\underline{X},\ \underline{L}'\bar{\underline{\beta}}) &= (\ \underline{b}' - \underline{\lambda}'X')\ X\underline{\lambda}\ \sigma^2 \\ &= (\ \underline{b}'X - \underline{\lambda}'X'X)\,\underline{\lambda}\ \sigma^2 \\ &= (\underline{L}' - \underline{L}')\,\underline{\lambda}\ \sigma^2 \\ &= 0. \end{aligned}$$

From the argument following Theorem 3.2 and Theorem 3.3 it follows that if $\underline{L}'\underline{\beta}$ is estimable, the unique best linear unbiased estimator of $\underline{L}'\underline{\beta}$ is $\underline{L}'\bar{\underline{\beta}}$, where $\bar{\underline{\beta}}$ is any solution of (3.17).

Theorem 3.6. If the linear parametric function $\phi_1 = \underline{L}_1'\underline{\beta}, \ldots, \phi_k = \underline{L}_k'\underline{\beta}$ are estimable, any linear combination of $\phi_1, \ldots, \phi_k$ is also estimable.

Proof. Let $\phi = \sum_{i=1}^{k} a_i\phi_i$, where $a_1, \ldots, a_k$ are real constants and let $\bar{\phi}_i$ be the unique linear unbiased estimator of ϕ_i, $i = 1,\ldots,k$. Then $\bar{\phi} = \sum_{i=1}^{k} a_i \bar{\phi}_i$ is obviously the unique unbiased linear estimator for ϕ. Thus ϕ is estimable.

Q.E.D.

It is easy to conclude that $\phi = \sum_{i=1}^{k} a_i \, L'_i \bar{\underline{\beta}}$, where $\bar{\underline{\beta}}$ is any solution of (3.17).

Theorem 3.7. All linear parametric functions in $\underline{\beta}$ are estimable if and only if X has full rank.

Proof. If X has full rank, the normal equations (3.11) have the unique solution $\underline{\beta} = (X'X)^{-1}X'\underline{x}$. Hence $\underline{L}'\bar{\underline{\beta}}$ is the unique estimator of $\underline{L}'\underline{\beta}$ for any given $\underline{L}$. Conversely if $\underline{L}'\underline{\beta}$ is estimable then by Theorem 3.2, $\underline{L} = X'X\underline{\lambda}$ for some $\underline{\lambda}$. Since $\underline{L}$ is arbitrary X'X must be of full rank.

Q.E.D.

If X is not of full rank some linear parametric functions do not admit of unbiased linear estimators and under the present set up nothing can be inferred about them. Linear parametric functions which are not estimable are said to be "confounded". In such situations we add linear restrictions on $\underline{\beta}$, so as to reduce the linear model to one of full rank. Such a reparametrization is mainly achieved by means of the transformation $\underline{\alpha} = A\underline{\beta}$, where A is a matrix of known elements such that each component of $\underline{\alpha}$ is an estimable function of $\underline{\beta}$. This will be discussed in more detail in the experimental design model.

Let r be the rank of X and $(X'X)^-$ denote the generalized inverse of X'X. Since $E(\underline{X}) = X\underline{\beta}$, all parametric functions $X\underline{\beta}$ are estimable and hence, by Theorem 3.3, $X\underline{\beta}$ admits of a unique least squares estimate $X\bar{\underline{\beta}}$, no matter what solution $\bar{\underline{\beta}}$ of (3.11) we take. Let us take in particular

$$\bar{\underline{\beta}} = (X'X)^- X'\underline{x} \, . \qquad (3.18)$$

From the identity

$$\begin{aligned} E(X\bar{\underline{\beta}}) &= E(X(X'X)^- X'\underline{X}) \\ &= X(X'X)^- X'X\underline{\beta} \\ &= X\underline{\beta} \end{aligned}$$

for all $\underline{\beta}$, it follows that we must have

$$X(X'X)^{-}X'X = X\ . \tag{3.19}$$

An estimator of σ^2 based on the unique least squares estimator $X\underline{\beta}$ of $X\underline{\beta}$ is given by

$$\begin{aligned}\bar{\sigma}^2 &= \frac{1}{(N-r)}\ (\underline{X}-X\underline{\beta})'\ (\underline{X}-X\underline{\beta})\\ &= \frac{1}{(N-r)}\ \underline{X}'(I-X(X'X)^{-}X')\ (I-X(X'X)^{-}X')\underline{X}\\ &= \frac{1}{(N-r)}\ \underline{X}'(I-X(X'X)^{-}X')\underline{X}\ ,\end{aligned} \tag{3.20}$$

as by (3.19),

$$\begin{aligned}&(I-X(X'X)^{-}X')(I-X(X'X)^{-}X')\\ &\quad = I-X(X'X)^{-}X'\ .\end{aligned} \tag{3.21}$$

In view of (3.21), the $N \times N$ matrix $I-X(X'X)^{-}X'$ is idempotent and its rank is equal to

$$\mathrm{tr}\ (I-X(X'X)^{-}X) = N-r\ .$$

Since

$$X'(I-X(X'X)^{-}X')X = 0,$$

we can write $\bar{\sigma}^2$ as

$$\bar{\sigma}^2 = \frac{1}{(N-r)}\ (\underline{X}-X\underline{\beta})'\ (I-X(X'X)^{-}X')(\underline{X}-X\underline{\beta})\ . \tag{3.22}$$

Lemma 3.8. $\bar{\sigma}^2$ is an unbiased estimator of σ^2.

Proof.

$$\begin{aligned}E(\bar{\sigma}^2) &= \frac{1}{(N-r)}\ \sigma^2\ \mathrm{tr}(I-X(X'X)^{-}X')\\ &= \frac{N-r}{N-r}\ \sigma^2 = \sigma^2\ .\end{aligned}$$

Q.E.D.

In what sense is $\bar{\sigma}^2$ an optimum estimator of σ^2 beyond being unbiased? This question was examined by Hsu (1938) and Rao (1952). Hsu considered the quadratic estimate of the form $\underline{X}'A\underline{X}$, which is unbiased for σ^2 and which has minimum variance among all quadratic unbiased estimators with variances independent of $\underline{\beta}$. Rao also considered unbiased quadratic estimators of the form $\underline{X}'A\underline{X}$ which have minimum variance among all unbiased nonnegative quadratic estimators. Under either condition it has been shown that $\bar{\sigma}^2$ has the minimum variance when the Pearsonian coefficient of Kurtosis of the distribution of each X_i is equal to 3. In the case when this coefficient is not equal to 3, $\bar{\sigma}^2$ has minimum variance when the coefficients of $X_1^2,\ldots,X_N^2$ in $\underline{X}'A\underline{X}$ are all equal.

If X'X has full rank, then,

$$\bar{\sigma}^2 = \frac{1}{(N-p)}\,\underline{X}'(I-X(X'X)^{-1}X')\underline{X}.$$

Let $\underline{L}'_1\underline{\beta},\underline{L}'_2\underline{\beta}$ be two estimable parametric functions and let $\underline{L}'_1\underline{\bar{\beta}}$, $\underline{L}'_2\underline{\bar{\beta}}$ be their least squares estimators.

Lemma 3.9.

$$\mathrm{Var}(\underline{L}'_1\underline{\bar{\beta}}) = \sigma^2 L'_1(X'X)^-\underline{L}_1\ ,$$

$$\mathrm{Cov}(\underline{L}'_1\underline{\bar{\beta}},\ \underline{L}'_2\underline{\bar{\beta}}) = \sigma^2\,\underline{L}'_1\,(X'X)^-\underline{L}_2\ .$$

Proof. Since $\underline{L}'_1\underline{\beta}$ is estimable, $\underline{L}_1 = (X'X)\underline{\lambda}$ for some $\underline{\lambda}$. By (3.18)

$$\begin{aligned}
&\mathrm{Var}(\underline{L}'_1\underline{\bar{\beta}})\\
&= \mathrm{Var}(\underline{L}'_1(X'X)^-X'\underline{X})\\
&= \sigma^2\,\underline{L}'_1(X'X)^-(X'X)(X'X)^-\underline{L}_1\\
&= \sigma^2\,\underline{\lambda}'(X'X)\ (X'X)^-(X'X)(X'X)^-\underline{L}_1\\
&= \sigma^2\,\underline{\lambda}'(X'X)\ (X'X)^-\underline{L}_1\\
&= \sigma^2\,\underline{L}'_1(X'X)^-\underline{L}_1\ .
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Cov}\ (\underline{L}'_1\underline{\bar{\beta}},\ L'_2\underline{\bar{\beta}})\\
&= \sigma^2\,\underline{L}'_1(X'X)^-(X'X)(X'X)^-\underline{L}_2\\
&= \sigma^2\,\underline{\lambda}'(X'X)\ (X'X)^-(X'X)(X'X)^-\underline{L}_2\\
&= \sigma^2\,\underline{\lambda}'\ X'X(X'X)^-\underline{L}_2\\
&= \sigma^2\,\underline{L}'_1\ (X'X)^-\underline{L}_2\ .
\end{aligned}$$

Q.E.D.

If (X'X) is of full rank, then all parametric functions are estimable. In particular, all components β_i of $\underline{\beta}$ are estimable and

$$\mathrm{Var}(\bar{\beta}_i) = \sigma^2 c_{ii},\quad \mathrm{Cov}(\bar{\beta}_i, \bar{\beta}_j) = \sigma^2 c_{ij} \tag{3.23}$$

where $C = (c_{ij}) = (X'X)^{-1}$.

3.2.2. Geometrical Approach to Least Squares: Let P be a projection operator (matrix) which projects vectors on the column space of X. By the definition of projection we get $PX = X$ and P is a symmetric idempotent matrix.

Let $\underline{b}'\underline{X}$ be an unbiased linear estimator of $\underline{L}'\underline{\beta}$. Since $E(\underline{b}'\underline{X}) = \underline{L}'\underline{\beta}$ for all $\underline{\beta}$, $\underline{b}'\underline{X} = \underline{L}'$. Consider $P\underline{b}$ and the linear estimator $(P\underline{b})'\underline{X} = \underline{b}'P\underline{X}$. Now

$$\begin{aligned} E(\underline{b}'P\underline{X}) &= \underline{b}'BX\underline{\beta} \\ &= \underline{b}'X\underline{\beta} \\ &= \underline{L}'\underline{\beta}\ , \end{aligned} \tag{3.24}$$

$$\begin{aligned} \mathrm{Var}(\underline{b}'X) &= \mathrm{Var}(\underline{b}'(I-P)\underline{X} + \underline{b}'P\underline{X}) \\ &= \mathrm{Var}(\underline{b}'(I-P)\underline{X}) + \mathrm{Var}(\underline{b}'P\underline{X}) \\ &\geq \mathrm{Var}(\underline{b}'P\underline{X})\ , \end{aligned} \tag{3.25}$$

as

$$\begin{aligned} \mathrm{Cov}&(\underline{b}'(I-P)\underline{X},\ \underline{b}'P\underline{X}) \\ &= \sigma^2\underline{b}'(I-P)P\underline{b} = \underline{0}, \quad \text{since P is idempotent.} \end{aligned}$$

Hence $\underline{b}'P\underline{X}$ is unbiased for $\underline{L}'\underline{\beta}$ and has smaller variance than $\underline{b}'\underline{X}$. If $\underline{\alpha}'\underline{X}$ is any other unbiased estimator of $\underline{L}'\underline{\beta}$ then $(\underline{b}' - \underline{\alpha}')\underline{X} = \underline{0}$, that is, $\underline{b} - \underline{\alpha}$ is orthogonal to the columns of X and therefore $P(\underline{b} - \underline{\alpha}) = \underline{0}$. So

$$\underline{\alpha}'P\underline{X} = \underline{b}'P\underline{X}\ .$$

Hence

$$\mathrm{Var}(\underline{\alpha}'\underline{X}) \geq \mathrm{Var}(\underline{\alpha}'P\underline{X}) = \mathrm{Var}(\underline{b}'P\underline{X})\ . \tag{3.26}$$

If $X\underline{\lambda}$ is the projection of the vector $\underline{b}$ on column space X then

$$X'(\underline{b} - X\underline{\lambda}) = \underline{0},$$

that is,

$$\underline{\lambda} = (X'X)^{-}X'\underline{b}\ ,$$
$$X\underline{\lambda} = X(X'X)^{-}X'\underline{b}\ .$$

So we can choose the projection matrix P as

$$P = X(X'X)^{-}X' .$$

Hence

$$\begin{aligned} P\underline{b} &= X(X'X)^{-}X'\underline{b} \\ &= X(X'X)^{-}X'\underline{L} , \end{aligned}$$

$$\begin{aligned} \underline{b}'P\underline{X} &= \underline{L}'(X'X)^{-}X'\underline{X} \\ &= \underline{L}'\underline{\bar{\beta}} . \end{aligned}$$

Thus $\underline{b}'P\underline{X}$ is the least squares estimator of $\underline{L}'\underline{\beta}$.

3.2.3. The Best Linear Unbiased Estimator and the Least Squares Estimator for Correlated Observations: Let $\underline{X} = (X_1,\ldots,X_N)'$ be an N-dimensional random vector with $E(\underline{X}) = X\underline{\beta}$, $\mathrm{Cov}(\underline{X}) = \sigma^2 V$, where V is a symmetric positive definite matrix. In this section we shall investigate conditions under which a best linear unbiased estimator coincides with the least squares estimator. This model is a slight generalization over the linear model (3.9) where V = I. We shall first consider the case where X has full rank and then we shall discuss the general case.

Since V is a positive definite matrix we can write

$$V = DD' , \tag{3.27}$$

where D is a nonsingular matrix. Writing $\underline{X}^* = D^{-1}\underline{X}$

$$E(\underline{X}^*) = D^{-1}X\underline{\beta} , \quad \mathrm{Cov}(\underline{X}^*) = \sigma^2 I. \tag{3.28}$$

From Theorem 3.5, the best linear unbiased estimator $\underline{\tilde{\beta}}$ of $\underline{\beta}$, assuming X has full rank, is given by

$$\underline{\tilde{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{X} , \tag{3.29}$$

with covariance matrix

$$\mathrm{Cov}(\underline{\tilde{\beta}}) = \sigma^2(X'V^{-1}X)^{-1} .$$

The least squares estimator $\underline{\bar{\beta}}$ of $\underline{\beta}$ is

$$\underline{\bar{\beta}} = (X'X)^{-1}X'\underline{X} \tag{3.30}$$

with $E(\underline{\bar{\beta}}) = \underline{\beta}$, $\mathrm{Cov}(\underline{\bar{\beta}}) = \sigma^2(X'X)^{-1}(X'VX)(X'X)^{-1}$.

Let us now investigate conditions of equivalence of $\underline{\tilde{\beta}}$ and $\underline{\bar{\beta}}$.

Magness and McGuire (1962) examined the simplest case $p = 1$, that is, $\underline{\beta}$ is a scalar quantity β, with the added restriction that $X'X = 1$. For simplicity they took $\mathrm{Var}(X_i) = 1$ for all i, so that $\sigma^2 V$ is the nonsingular correlation matrix R of $\underline{X}$. Let $\theta_1, \ldots, \theta_N$ be the characteristic roots of R and let $\underline{\alpha}_1, \ldots, \underline{\alpha}_N$ be the corresponding normalized characteristic vectors. So

$$R\underline{\alpha}_i = \theta_i \underline{\alpha}_i , \quad i = 1, \ldots, N ,$$

$$\underline{\alpha}_i{}' \underline{\alpha}_j = 0, \quad i \neq j .$$

Since the set of N orthonormal characteristic N-vectors $\underline{\alpha}_1, \ldots, \underline{\alpha}_N$ constitutes a basis to the N-dimensional vector space, we can write

$$X = \sum_{i=1}^{N} C_i \underline{\alpha}_i$$

where $C_1, \ldots, C_N$ are scalar constants. From the condition that $\underline{X}'\underline{X} = 1$ and $\underline{\alpha}_i$'s are orthonormal vectors, we get, $\sum_{i=1}^{N} C_i^2 = 1$. Thus for the best linear unbiased estimator $\tilde{\beta}$,

$$\mathrm{Var}(\tilde{\beta}) = (X'R^{-1}X)^{-1}$$

$$= [(\sum_{i=1}^{N} C_i \underline{\alpha}_i)' R^{-1} (\sum_{i=1}^{N} C_i \underline{\alpha}_i)]^{-1} .$$

Since $\underline{\alpha}_i$ are also the characteristic vectors of R^{-1} ~~corresponding~~ corresponding to the characteristic roots θ_i^{-1}, $i = 1, \ldots, N$, that is,

$R^{-1}\underline{\alpha}_i = \theta_i^{-1} \underline{\alpha}_i$, we have

$$\sum_{i=1}^{N} C_i R^{-1} \underline{\alpha}_i = \sum_{i=1}^{N} \left(\frac{C_i}{\theta_i}\right) \underline{\alpha}_i .$$

So

$$\mathrm{Var}(\tilde{\beta}) = \left[\sum_{i=1}^{N} \frac{C_i^2}{\theta_i}\right]^{-1} .$$

For the least squares estimator $\hat{\beta}$ of β

$$\mathrm{Var}(\hat{\beta}) = X'RX = \sum_{i=1}^{N} C_i^2 \theta_i .$$

Generally $\mathrm{Var}(\tilde{\beta}) \leq \mathrm{Var}(\hat{\beta})$. The condition under which equality holds is given in the following theorem which is due to Magness and

McGuire (1962).

Theorem 3.10. $\operatorname{Var}(\tilde{\beta}) = \operatorname{Var}(\bar{\beta})$ if and only if X is a characteristic vector of R.

Proof. If λ is the characteristic root of R corresponding to the vector X, then,

$$R^{-1}X = \frac{1}{\lambda}X, \quad RX = \lambda X.$$

So

$$\tilde{\beta} = X'\underline{X} = \bar{\beta}$$

and hence $\operatorname{Var}(\bar{\beta}) = \operatorname{Var}(\tilde{\beta}) = \lambda.$

Conversely, suppose that

$$\operatorname{Var}(\tilde{\beta}) = \operatorname{Var}(\bar{\beta}) = \lambda.$$

Let $X^{\perp}$ be any unit vector perpendicular to X.

Then

$$RX = \lambda X + \mu X^{\perp},$$

or

$$\lambda X = RX - \mu X^{\perp},$$

where λ, μ are scalar constants. So

$$X'RX = \lambda X'X = \lambda > 0.$$

Now, $X = \lambda R^{-1}X + \mu R^{-1}X^{\perp}$.

So

$$R^{-1}X = \frac{1}{\lambda}X - \frac{\mu}{\lambda} R^{-1} X^{\perp},$$

$$X'R^{-1}X = \frac{1}{\lambda} - \frac{\mu}{\lambda} X'R^{-1}X^{\perp}.$$

Hence,

$$\begin{aligned}\lambda = \operatorname{Var}(\tilde{\beta}) &= (X'R^{-1}X)^{-1} \\ &= \lambda(1-\mu X'R^{-1}X^{\perp})^{-1}.\end{aligned} \tag{3.31}$$

Since $\lambda > 0$ we conclude from (3.31) that either $\mu = 0$ or $X'R^{-1}X^{\perp} = 0$. But $X^{\perp}$ perpendicular to X implies that $\mu = 0$. Hence $RX = \lambda X$ and $R^{-1}X = X/\lambda$, which implies that X is a characteristic vector of R corresponding to the characteristic root λ.

Q.E.D.

In the general case with $E(\underline{X}) = X\underline{\beta}$, $\operatorname{Cov}(\underline{X}) = \sigma^2 V$, where V is a symmetric nonnegative definite matrix and X is not of full rank, Zyskind (1967) proved the following theorem.

Theorem 3.11. Let r be the rank of X and let $\underline{L}'\underline{\beta}$ be estimable. The best linear unbiased estimator of $\underline{L}'\underline{\beta}$ coincides with its least squares estimator if and only if there exists a subset of r characteristic vectors of V which constitutes a basis of the subspace $C(X)$ spanned by the column vectors of X.

Proof. Suppose that V has r orthogonal characteristic vectors which form the basis of the linear subspace spanned by the rows of X. Let M_0 be the $r \times N$ matrix whose rows are these r orthogonal characteristic vectors. The other $N - r$ characteristic vectors of V form a basis of the orthogonal complement of $C(X)$ and let M_1 be the $(N - r) \times N$ matrix whose row vectors are these $N - r$ characteristic vectors. Obviously $M_1X = 0$.

Let $M = \begin{pmatrix} M_0 \\ M_1 \end{pmatrix}$ and let $\underline{Y} = M\underline{X}$. Then

$$E(\underline{Y}) = (M_0X\underline{\beta}),$$

$$\mathrm{Cov}(\underline{Y}) = \sigma^2 MVM' = \sigma^2\Delta,$$

where Δ is a diagonal matrix whose diagonal elements are the characteristic roots of V. If the rank of V is less than N then some of these diagonal elements are zeros and the others are positive. Thus the best linear unbiased estimator of any estimable parametric function $\underline{L}'\underline{\beta}$ should only be a function of $M_0\underline{X}$.

Let $\underline{b}'\underline{X}$ be any unbiased estimator of $\underline{L}'\underline{\beta}$ and let $\underline{b} = \underline{a} + \underline{a}^{\perp}$, where $\underline{a}^{\perp}$ is the projection of $\underline{b}$ on $C(X)$ and $\underline{a}$ is the orthogonal complement of $\underline{a}$. Now

$$\begin{aligned} E(\underline{b}'\underline{X}) &= E(\underline{a}'\underline{X}) + (\underline{a}^{\perp})'\underline{X}) \\ &= E(\underline{a}'\underline{X}) + (\underline{a}^{\perp})'X\underline{\beta} \\ &= E(\underline{a}'\underline{X}). \end{aligned}$$

Thus $\underline{a}'\underline{X}$ is also an unbiased estimator of $\underline{L}'\underline{\beta}$. Again, since

$$M\underline{a} = \begin{pmatrix} M_0 \\ M_1 \end{pmatrix} \underline{a}$$

$$= \begin{pmatrix} M_0\underline{a} \\ \underline{0} \end{pmatrix} ,$$

$$M\underline{a}^{\perp} = \begin{pmatrix} \underline{0} \\ M_1\underline{a}^{\perp} \end{pmatrix} ,$$

$$\begin{aligned} \mathrm{Var}(\underline{b}'\underline{X}) &= \sigma^2\underline{b}'\,V\,\underline{b} \\ &= \sigma^2(\underline{a}+\underline{a}^{\perp})'V(\underline{a}+\underline{a}^{\perp}) \\ &= \sigma^2(\underline{a}+\underline{a}^{\perp})'M'DM(\underline{a}+\underline{a}^{\perp}) \\ &= \sigma^2[\underline{a}'M_0'D_0M\underline{a} + (M_1\underline{a}^{\perp})'D_1(M_1\underline{a}^{\perp})] \\ &\geq \sigma^2\;\underline{a}'M_0'D_0M_0\underline{a} \\ &= \sigma^2\;\underline{a}'V\underline{a} \\ &= \mathrm{Var}(\underline{a}'\underline{X}) , \end{aligned}$$

where

$$D = \begin{pmatrix} D_0 & 0 \\ 0 & D_1 \end{pmatrix} .$$

Thus the best linear unbiased estimator of $\underline{L}'\underline{\beta}$ is $\underline{a}'\underline{X}$, in which $\underline{a}$ belongs to $C(X)$. Moreover since M_0 is a basis of $C(X)$

$$\begin{aligned} \underline{a}'\underline{X} &= \underline{a}'\;M_0'\;M_0\underline{X} \\ &= E'\;(M_0\underline{X}) , \end{aligned}$$

where $E = M_0\underline{a}$. Finally since $C(X) = C(M_0X)$, the best linear unbiased estimator of $L'\underline{\beta}$ is the least squares estimator which is the unique unbiased estimator of $\underline{L}'\underline{\beta}$ for which $\underline{L}\,\varepsilon\,C(M_0X)$.

Conversely, suppose that the best linear unbiased estimator of the estimable parametric function $\underline{L}'\underline{\beta}$ coincides with the least squares estimator. Let $M = \begin{pmatrix} M_0 \\ M_1 \end{pmatrix}$ be an orthogonal matrix such that $M_1X = 0$. Let $\underline{Y} = M\underline{X}$. Then

$$E(\underline{Y}) = \begin{pmatrix} M_0X\underline{\beta} \\ \underline{0} \end{pmatrix} .$$

Since $M_0'M_0 = I$, $M_0\underline{X}$ is the best linear unbiased estimator of

$M_0X\underline{\beta}$. This implies that Cov($\underline{Y}$) is of the form

$$\sigma^2 \begin{pmatrix} W_0 & 0 \\ 0 & W_1 \end{pmatrix},$$

otherwise the least squares estimator of $M_0X\underline{\beta}$ is best linear unbiased. Let R_0, R_1 be orthogonal matrices such that

$$R_0W_0R_0' = \Delta_0 ,$$

$$R_1W_1R_1' = \Delta_1 ,$$

where Δ_0, Δ_1 are diagonal matrices. Let

$$P = \begin{pmatrix} R_0 & 0 \\ 0 & R_1 \end{pmatrix} \begin{pmatrix} M_0 \\ M_1 \end{pmatrix}$$

$$= \begin{pmatrix} R_0 & M_0 \\ R_1 & M_1 \end{pmatrix}$$

Then

$$PVP' = \begin{pmatrix} D_0 & 0 \\ 0 & D_1 \end{pmatrix}$$

Thus the column vectors of P form a set of N characteristic vectors of V. The row vectors of R_0M_0 are linear combinations of those of M_0 and similarly those of R_1M_1. Hence R_0M_C, R_1M_1 form respectively orthonormal bases of C(X) and its orthogonal complement.

Q.E.D.

3.2.4. Least Squares Estimator Subject to Linear Restrictions on Parameters: Let $\underline{X}$ have an N-variate normal distribution with mean $X\underline{\beta}$ and covariance matrix σ^2I and let $\underline{\beta}$ be subject to consistent linear restriction $C'\underline{\beta} = \underline{\theta}$ (given) with rank C = p-r. Suppose that a general solution of $C'\underline{\beta} = \theta$ is $\underline{\beta}^0 + B\underline{\alpha}$, where $\underline{\beta}^0$ is a particular solution of $C'\underline{\beta} = \underline{\theta}$, is B is a p × r matrix of full rank r such that C'B = 0, and $\underline{\alpha}$ an arbitrary vector of dimension r. Then

$$\begin{aligned} E(\underline{X}) &= X\underline{\beta} \\ &= X(\underline{\beta}^{o} + B\underline{\alpha}) , \end{aligned} \tag{3.32}$$

or

$$E(\underline{X} - X\underline{\beta}^{o}) = XB\underline{\alpha} .$$

Since $\underline{Z} = \underline{X} - X\underline{\beta}^{o}$ has an N-dimensional normal distribution with mean $XB\underline{\alpha}$ and covariance matrix $\sigma^2 I$, the problem of estimating $\underline{\beta}, \sigma^2$ is reduced to that of considering new variable $\underline{Z}$ and estimating new parameters $\underline{\alpha}, \sigma^2$.

For the estimation of the linear parametric function $\underline{L}'\underline{\beta} = \underline{L}'(\underline{\beta}^{o} + B\underline{\alpha})$ and σ^2, we need to know the rank of XB. Since $C'B = 0$, the rank of the augmented matrix $(X|C')$ satisfies

$$\text{rank}(X|C') = \text{rank}(XB) + \text{rank}(C) \tag{3.33}$$

that is, the rank of XB can be obtained by studying the matrices X,C without actually computing B and XB.

An alternative approach which avoids such direct verification of ranks is as follows: Consider the problem of finding a linear function

$$\underline{b}'\underline{X} + \underline{d} ,$$

such that

$$\text{Var}(\underline{b}'\underline{X} + \underline{d}) = \sigma^2 \underline{b}'\underline{b} \tag{3.34}$$

is minimum for all choices of $\underline{b}, \underline{d}$, subject to the condition that

$$\begin{aligned} E(\underline{b}'\underline{X} + \underline{d}) &= \underline{b}'X\underline{\beta} + \underline{d} \\ &= \underline{L}'\underline{\beta} , \end{aligned} \tag{3.35}$$

and $C'\underline{\beta} = \underline{\theta}$.

Write $\underline{d} = M'\theta$. Noting that $C'\underline{\beta} = \underline{\theta}$, the problem is reduced to minimizing $\underline{b}'\underline{b}$ subject to the restriction that

$$\underline{b}'X + M'C' = \underline{L}' , \tag{3.36}$$

or $(X' \;\; C) \begin{pmatrix} \underline{b} \\ M \end{pmatrix} = \underline{L}$.

Theorem 3.12. Let

$$\begin{pmatrix} X'X , & C' \\ C , & 0 \end{pmatrix}^{-} = \begin{pmatrix} C_1 , & C_2 \\ C_3 , & C_4 \end{pmatrix} \tag{3.37}$$

Then,

$$\inf_{\underline{b}'X + M'C' = \underline{L}'} \underline{b}'\underline{b} = (\underline{b}^*)'\underline{b}^* = \underline{L}'C_1\underline{L}$$

where $\underline{b}^* = XC_1'\underline{L}$, and the corresponding value of M is $M^* = C_2'\underline{L}$.

Proof. From (3.36) we conclude that there exist column vectors $\underline{E},\underline{F}$ such that

$$\underline{L} = (X'X)\underline{E} + C'\underline{F}$$

$$0 = C\underline{E} \cdot$$

With

$$\underline{b}^* = XC_1'\underline{L}, \quad M^* = C_2'L ,$$

we get

$$(X'|C)\begin{pmatrix} XC_1' & \underline{L} \\ C_2' & \underline{L} \end{pmatrix} = \underline{L}$$

and

$$(\underline{b}^*)'(\underline{b}^*) = \underline{L}'C_1X'XC_1'\underline{L} = \underline{L}'C_1\underline{L}$$

Note that if A^- is a generalized inverse of A then $A^-AA^- = A^-$. Furthermore for any $\underline{b}$ satisfying (3.36) $(\underline{b}^*)'\underline{b} = \underline{L}'C_1\underline{L}$. Now by applying Cauchy-Schwartz inequality we obtain

$$[(\underline{b}^*)'\underline{b}]^2 \le ((\underline{b}^*)'\underline{b}^*)\ (\underline{b}'\underline{b}) ,$$

or

$$(\underline{b}'\underline{b}) \ge \frac{[(\underline{b}^*)'\underline{b}]^2}{(b^*)'\underline{b}^*} = \underline{L}'C_1\underline{L} .$$

Q.E.D.

Thus the minimum variance unbiased estimator of $\underline{L}'\underline{\beta}$ is

$$(\underline{b}^*)'\underline{X} + (M^*)'\underline{e} = \underline{L}'\ (C_1Y'\underline{X} + C_2\underline{e}) , \tag{3.38}$$

which can also be written as $\underline{L}'\hat{\underline{\beta}}$ where $\hat{\underline{\beta}}$ is a solution of

$$X'X\underline{\beta} + C\underline{\gamma} = X'\underline{X} ,$$

subject to the restriction $C'\underline{\beta} = \underline{e}$. (3.39)

Observe that (3.39) can be obtained by minimizing $(\underline{X} - X\underline{\beta})'(\underline{X} - X\underline{\beta})$ subject to the restriction $C'\underline{\beta} = \underline{e}$, using the Lagrange multiplier $\underline{\gamma}$. Now since

$$\underline{L}'\hat{\underline{\beta}} = \underline{L}'C_1X'\underline{X} + \underline{L}'C_2\underline{\theta} \quad ,$$

from (3.37)

$$\begin{aligned} \mathrm{Var}(\underline{L}'\hat{\underline{\beta}}) &= \sigma^2\underline{L}'C_1X'XC_1\underline{L} \\ &= \sigma^2\underline{L}'C_1\underline{L} \quad . \end{aligned} \qquad (3.40)$$

Similarly for linear parametric functions $\underline{L}_1'\underline{\beta}$, $\underline{L}_2'\underline{\beta}$,

$$\begin{aligned} \mathrm{Cov}(\underline{L}_1'\hat{\underline{\beta}}, \underline{L}_2'\hat{\underline{\beta}}) &= \sigma^2\underline{L}_1'C_1\underline{L}_2 \\ &= \sigma^2\underline{L}_2'C_1\underline{L}_1 \quad . \end{aligned} \qquad (3.41)$$

In the case where the matrix C_1 is not computed but it is known that

$$\underline{L}_1'\hat{\underline{\beta}} = \underline{\gamma}'X'\underline{X} \ , \ \underline{L}_2'\hat{\underline{\beta}} = \hat{\underline{\delta}}'X'\underline{X} \ , \qquad (3.42)$$

satisfying the condition that

$$C'\underline{\gamma} = C'\underline{\delta} = \underline{0} \ , \qquad (3.43)$$

then

$$\begin{aligned} \mathrm{Var}(\underline{L}_1'\hat{\underline{\beta}}) &= \sigma^2\underline{L}_1'\hat{\underline{\gamma}}, \ \mathrm{Var}(\underline{L}_2'\hat{\underline{\beta}}) = \sigma^2\underline{L}_2'\underline{\delta} \\ \mathrm{Cov}(\underline{L}_1'\hat{\underline{\beta}}, \underline{L}_2'\hat{\underline{\beta}}) &= \sigma^2\underline{L}_1'\underline{\delta} = \sigma^2\underline{L}_2'\underline{\gamma} \ . \end{aligned} \qquad (3.44)$$

If $C'\underline{\gamma} = C'\underline{\delta} = \underline{0}$ is not satisfied then

$$\begin{aligned} \mathrm{Var}(\underline{L}_1'\hat{\underline{\beta}}) &= \sigma^2\underline{\gamma}'X'X\underline{\gamma} \ , \ \mathrm{Var}(\underline{L}_2'\hat{\underline{\beta}}) = \sigma^2\underline{\delta}'X'X\underline{\delta} \\ \mathrm{Cov}(\underline{L}_1'\hat{\underline{\beta}}, \underline{L}_2'\hat{\underline{\beta}}) &= \sigma^2\underline{\gamma}'X'X\underline{\delta} \ . \end{aligned} \qquad (3.45)$$

3.2.5. Least Squares in Random Effect Models: Consider the linear model (3.9) where $\underline{\beta}$ is a random vector with mean $\underline{\mu}$ and covariance matrix T. We shall be interested here in finding the best linear estimator (predictor) of the random variable $\underline{L}'\underline{\beta}$. Since the conditional covariance of $\underline{X}$ given $\underline{\beta}$ is

$$\mathrm{Cov}(\underline{X}|\underline{\beta}) = \sigma^2 I \ ,$$

we get

$$\begin{aligned} \mathrm{Cov}(\underline{X}) &= E(\mathrm{Cov}(\underline{X}|\underline{\beta})) + \mathrm{Cov}(E(\underline{X}|\underline{\beta})) \\ &= \sigma^2 I + \mathrm{Cov}(X|\underline{\beta}) \\ &= \sigma^2 I + XTX' \ , \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}(\underline{X}, \underline{L}'\underline{\beta}) &= E(\mathrm{Cov}(\underline{X}, \underline{L}'\underline{\beta}|\underline{\beta})) + \mathrm{Cov}(E(\underline{X}|\underline{\beta}), \underline{L}'\underline{\beta}) \\ &= 0 + XT\underline{L}. \end{aligned}$$

To find the best linear estimator of $\underline{L}'\underline{\beta}$ we look for a linear function $a + \underline{b}'\underline{X}$ satisfying

$$E(a + \underline{b}'\underline{X} - \underline{L}'\underline{\beta}) = 0 , \qquad (3.46)$$

and

$$\mathrm{Var}(a + \underline{b}'\underline{X} - \underline{L}'\underline{\beta}) \text{ is minimum} \qquad (3.47)$$

for all choices of $a, \underline{b}$.

Theorem 3.13. If $\underline{\mu}$ is unknown, the values $\underline{b}^*$, a^* of $\underline{b}, a$ satisfying (3.46) and (3.47) are given by

$$\underline{b}^* = X(X'X)^{-1}\underline{L} , \quad a^* = 0 \qquad (3.48)$$

provided there exists a $\underline{b}$ such that $X'\underline{b} = \underline{L}$. Also the prediction variance, that is, $\mathrm{Var}((\underline{b}^*)'X - \underline{L}'\underline{\beta}) = \underline{L}'(X'X)^{-1}\underline{L}$.

Note. If $(X'X)^{-1}$ does not exist it is replaced by a generalized inverse $(X'X)^-$ of $X'X$.

Proof. Since

$$X'\underline{b}^* = (X'X)(X'X)^{-1}\underline{L} = \underline{L}$$

we get

$$\begin{aligned} E((\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta}) &= (\underline{b}^*)'X\underline{\mu} - \underline{L}'\underline{\mu} \\ &= \underline{L}'\underline{\mu} - \underline{L}'\underline{\mu} \\ &= 0 . \end{aligned}$$

Since, for any other choices $a, \underline{b}$ satisfying (3.46) and the condition that $X'\underline{b} = \underline{L}$,

$$\begin{aligned} &\mathrm{Cov}(\underline{L}'\underline{\beta} - (\underline{b}^*)'\underline{X},\ a + (\underline{b} - \underline{b}^*)'\underline{X}) \\ &= \mathrm{Cov}(\underline{L}'\underline{\beta} - (\underline{b}^*)'\underline{X},\ (\underline{b} - \underline{b}^*)'\underline{X}) \\ &= (\underline{b} - \underline{b}^*)'XT\underline{L} - \sigma^2(\underline{b} - \underline{b}^*)'b^* \\ &= (\underline{b}'X - (\underline{b}^*)'X)T\underline{L} - \sigma^2(\underline{b} - \underline{b}^*)'X(X'X)^{-1}\underline{L} \\ &= 0 - 0 \\ &= 0, \end{aligned}$$

we get

$$\begin{aligned} &\mathrm{Var}(a + \underline{b}'\underline{X} - \underline{L}'\underline{\beta}) \\ &= \mathrm{Var}((\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta} + a + (\underline{b} - \underline{b}^*)'\underline{X}) . \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Var}((\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta}) + \mathrm{Var}(a + (\underline{b} - \underline{b}^*)'\underline{X}) \\
&= \mathrm{Var}((\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta}) + \mathrm{Var}((\underline{b} - \underline{b}^*)'\underline{X}) \\
&\geq \mathrm{Var}((\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta}) \\
&= \sigma^2(\underline{b}^*)'(\underline{b}^*) \\
&= \sigma^2\underline{L}'(X'X)^{-1}\underline{L} \ .
\end{aligned}
$$

Q.E.D.

It is interesting to note that the best estimator and the prediction variance are the same as the least squares estimator and its variance respectively, when $\underline{\beta}$ is taken to be a fixed parameter .

In case $\underline{\mu}$ is known the optimum choices of $\underline{b}$ and $\underline{a}$ are

$$
\begin{aligned}
\underline{b}^* &= (\sigma^2 I + XTX')^{-1}XT\underline{L} \\
\underline{a}^* &= \underline{\mu}'\underline{L} - \underline{\mu}'X'\underline{b}^*
\end{aligned}
\tag{3.49}
$$

and the prediction variance is

$$
\begin{aligned}
V(a^* + (\underline{b}^*)'\underline{X} - \underline{L}'\underline{\beta}) &= \underline{L}'T\underline{L} - \underline{L}'TX'\underline{b}^* \\
&= \underline{L}'(T^{-1} + X'(\sigma^2)^{-1}X)^{-1}\underline{L} \ .
\end{aligned}
\tag{3.50}
$$

3.2.6. Maximum Likelihood and Least Squares: For testing hypotheses about linear parametric functions and finding the distributions of least squares estimators of different parameters we need now to specify the distributions of $X_1,\dots,X_N$. We shall assume throughout the remaining part of the chapter that $X_1,\dots,X_N$ are independently distributed normal variables with means and variances given by (3.8). The likelihood of the observations $x_1,\dots,x_N$ is then given by

$$
L(\underline{\beta},\sigma^2) = (2\pi)^{-\frac{1}{2}N}(\sigma^2)^{-\frac{1}{2}N} \exp\{-\frac{1}{2\sigma^2}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})\} \ . \tag{3.51}
$$

Taking logarithms we get

$$
\log L(\underline{\beta},\sigma^2) = -\tfrac{1}{2}N\log(2\pi) - \tfrac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta}) \ . \tag{3.52}
$$

Differentiating (3.52) with respect to the component β_k of $\underline{\beta}$ and σ^2, the maximum likelihood estimates $\hat{\underline{\beta}}$ of $\underline{\beta}$ and $\hat{\sigma}^2$ of σ^2 are given by

$$(X'X)\hat{\underline{\beta}} = X'\underline{x}\,, \tag{3.53}$$

$$\hat{\sigma}^2 = \frac{1}{N}(\underline{x}-X\hat{\underline{\beta}})'(\underline{x}-X\hat{\underline{\beta}})$$

$$= \frac{1}{N}\underline{x}'(I-X(X'X)^{-}X')\underline{x}\,.$$

Hence the maximum likelihood estimate of $\underline{\beta}$ under the assumption of normal distribution is the same as its least squares estimate.

If X has full rank, (X'X) has the unique inverse $(X'X)^{-1}$ and

$$E(\hat{\underline{\beta}}) = E((X'X)^{-1}X'X) = (X'X)^{-1}X'X\underline{\beta} = \underline{\beta}\,.$$

In the general case $\hat{\underline{\beta}}$ is not unique and $E(\hat{\underline{\beta}}) \neq \underline{\beta}$, and from lemma 3.8

$$E(\hat{\sigma}^2) = \frac{N-r}{N}\sigma^2\,.$$

__Theorem 3.14__. Under the assumption of normal distributions for $\underline{x}_1,\ldots,\underline{x}_N$, the maximum likelihood estimator (also the least squares estimator) $\underline{L}'\hat{\underline{\beta}}$ of the estimable linear parametric function $\underline{L}'\underline{\beta}$ is independent of $\hat{\sigma}^2$ and $\underline{L}'\hat{\underline{\beta}}$ has normal distribution with mean $\underline{L}'\underline{\beta}$ and variance $\underline{L}'(X'X)^{-}\underline{L}$, $N\hat{\sigma}^2/\sigma^2$ has the chisquare distribution with N-r degrees of freedom.

__Proof__. Let $A = I-X(X'X)^{-}X'$, $B = \underline{L}'(X'X)^{-}X'$. Then $\underline{L}'\hat{\underline{\beta}}=B\underline{X}$ and $N\hat{\sigma}^2 = (\underline{X}-X\underline{\beta})'(I-X(X'X)^{-}X')(X-X\underline{\beta})$. Using (3.19) we have

$$BA = \underline{L}'(X'X)^{-}(X'-(X'X)(X'X)^{-}X') = 0\,.$$

Hence by an argument similar to that of Theorem 2.9 we conclude that $\underline{L}'\hat{\underline{\beta}}$, $N\hat{\sigma}^2$ are independent. As $\underline{X}$ has an N-variate normal distribution with mean $X\underline{\beta}$ and covariance matrix $\sigma^2 I$ and $\underline{L}'\underline{\beta}$ is estimable, using Lemma 3.9, we conclude that $\underline{L}'\hat{\underline{\beta}} = B\underline{X}$ is normally distributed with mean $\underline{L}'\underline{\beta}$ and variance $\underline{L}'(X'X)^{-}\underline{L}$. Furthermore, since A is idempotent of rank N-r, by Theorem 2.7, $N\hat{\sigma}^2/\sigma^2$ is distributed as chisquare with N-r degrees of freedom.

Q.E.D.

__Corollary 1__. If X has full rank p, then $\hat{\underline{\beta}}$, $N\hat{\sigma}^2$ are independent and $\hat{\underline{\beta}}$ has the p-variate normal distribution with mean $\underline{\beta}$ and covariance matrix $\sigma^2(X'X)^{-1}$, $N\hat{\sigma}^2/\sigma^2$ is distributed as chisquare with N-p degrees of freedom.

Proof. Since X has full rank p, then p × p matrix X'X has the unique inverse $(X'X)^{-1}$ and

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{X} .$$

so

$$\begin{aligned} E(\underline{\hat{\beta}}) &= (X'X)^{-1}(X'X)\,\underline{\beta} \\ &= \underline{\beta}\,, \end{aligned}$$

$$\begin{aligned} \mathrm{cov}(\underline{\hat{\beta}}) &= E(\underline{\hat{\beta}} - \underline{\beta})\ (\underline{\hat{\beta}} - \underline{\beta})' \\ &= E((X'X)^{-1}X'\underline{X} - \underline{\beta})((X'X)^{-1}X'\underline{X} - \underline{\beta})' \\ &= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} . \end{aligned}$$

Since $(X'X)^{-1}X'$ is a p × N matrix of rank p $(p < N)$ and $\underline{X}$ has the N-variate normal distribution with mean $X\underline{\beta}$ and covariance $\sigma^2 I$, it follows from Theorem 2.5 that $\underline{\hat{\beta}}$ has a p-variate normal distribution with mean $\underline{\beta}$ and covariance matrix $\sigma^2 (X'X)^{-1}$.

In this case

$$\frac{N\hat{\sigma}^2}{\sigma^2} = \left(\frac{\underline{X} - X\underline{\beta}}{\sigma}\right)'\ \left(I - X(X'X)^{-1}\left(\frac{\underline{X} - X\underline{\beta}}{\sigma}\right)\right) . \tag{3.54}$$

Since $I - X(X'X)^{-1}X'$ is an idempotent matrix of rank N-p, $N\hat{\sigma}^2/\sigma^2$ is distributed as a chisquare with N-p degrees of freedom. The independence of $\underline{\hat{\beta}}$, $N\hat{\sigma}^2$ follows from the fact that $X'(I - X(X'X)^{-1}X') = 0$.

Q.E.D.

Corollary 2. If X has full rank p ,

(a) then $\hat{\beta}_i$ is normally distributed with mean β_i and

$$\mathrm{Var}(\beta_i) = c_{ii}\ \sigma^2 ,$$

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\ \sigma^2 ,$$

where $(X'X)^{-1} = C = (c_{ij})$.

(b) For any linear parametric function $\underline{L}'\underline{\beta}$, the maximum likelihood estimator $\underline{L}'\underline{\hat{\beta}}$ is normally distributed with mean $\underline{L}'\underline{\beta}$ and variance $\sigma^2 \underline{L}'(X'X)^{-1}\underline{L}$.

The proof is obvious.

Example 3.3. Let $Y_1,\dots,Y_N$ be independently distributed normal random variables with

$$E(Y_i) = \alpha + \beta x_i ,$$
$$\mathrm{Var}(Y_i) = \sigma^2 ,$$

$i = 1,\dots,N$, where $x_1,\dots,x_N$ are known constants (at least two of which are distinct) and α, β, σ^2 are unknown parameters. Here

$$X = \begin{pmatrix} 1 , & x_1 \\ \cdot , & \cdot \\ \cdot , & \cdot \\ 1 , & x_N \end{pmatrix} X'X = \begin{pmatrix} N , & N\bar{x} \\ & \\ N\bar{x} , & \sum_{i=1}^{N} x_i^2 \end{pmatrix},$$

$$(X'X)^{-1} = \frac{1}{N \sum_{i=1}^{N} (x_i-\bar{x})^2} \begin{pmatrix} \sum_{i=1}^{N} x_i^2 , & -N\bar{x} \\ & \\ -N\bar{x} , & N \end{pmatrix}$$

where $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$. The maximum likelihood estimates of α, β, σ^2 are

$$\hat{\beta} = \frac{\sum_{i=1}^{N} (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} , \qquad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ \sum_{i=1}^{N} (y_i-\bar{y})^2 - \frac{\left(\sum_{i=1}^{N} (x_i - \bar{x}) \; y_i \right)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right\} .$$

For the corresponding estimators

$$E(\hat{\beta}) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})\, E(Y_i - \bar{Y})}{\Sigma (x_i - \bar{x})^2} = \beta ,$$

$$E(\hat{\alpha}) = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha,$$

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad \mathrm{Var}(\hat{\alpha}) = \frac{\sigma^2}{N} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2},$$

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = - \frac{\sigma^2 \bar{x}}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad E(\hat{\sigma}^2) = \frac{(N-2)\sigma^2}{N} .$$

Example 3.4. Consider Example 3.3 with

$$E(Y_i) = \alpha + \beta(x_i - \bar{x})$$

$$\mathrm{Var}(Y_i) = \sigma^2, \; i = 1, \ldots, N .$$

It is easy to check that the maximum likelihood estimates (which are the same as the least squares estimates) of α, β are

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2},$$

$$\hat{\sigma}^2 = \frac{1}{N}\left(\sum_{i=1}^{N} (y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x}) \right]^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right)$$

The corresponding estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ satisfy

$$E(\hat{\alpha}) = \alpha, \quad E(\hat{\beta}) = \beta,$$

$$E(\hat{\sigma}^2) = \frac{N-2}{N}\sigma^2, \quad \mathrm{Var}(\hat{\alpha}) = \frac{\sigma^2}{N},$$

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = 0 .$$

3.3. PERCENT POINTS AND CONFIDENCE INTERVAL

The upper α-percent point of the distribution of a random variable X is the value $x_{1-\alpha}$ of the random variable such that $P(X \ge x_{1-\alpha}) = \alpha$. The lower α-percent point of the distribution is the value x_α such that $P(X \le x_\alpha) = \alpha$. Obviously if the distribution is symmetric about the origin then $x_\alpha = -x_{1-\alpha}$. The values $\chi^2_{1-\alpha,n}$, $\chi^2_{\alpha,n}$ will denote the upper and lower α-percent points of the chisquare distribution with n degrees of freedom respectively. Similarly we will use the notations $t_{1-\alpha,n}$ $t_{\alpha,n}$ and $F_{1-\alpha,n_1,n_2}$, F_{α,n_1,n_2} to denote the upper and lower α-percent points of the t-distribution (central) with n degrees of freedom and of the F-distribution with n_1, n_2 degrees of freedom respectively. For different α these values can be found from statistical tables.

From the F-distribution (see Giri, 1974), it is easy to verify that

$$F_{1-\alpha,n_1,n_2} = (F_{\alpha,n_2,n_1})^{-1} \tag{3.56}$$

The percent points of the chisquare and the t-distribution can also be found from the F-distribution by using the relations

$$\chi^2_{\alpha,n} = n\, F_{\alpha,n,\infty} , \tag{3.57}$$

$$t_{\alpha/2,n} = (F_{\alpha,1,n})^{\frac{1}{2}} \tag{3.58}$$

Let $\underline{X}$ be distributed as a normal N-vector with mean and covariance as given in (3.9). To find the confidence interval of β_k, assuming that it is estimable we use the fact (3.54) that $\hat{\beta}_k$ is normally distributed with mean β_k and variance $C_{kk}\sigma^2$, and

$$\frac{N\hat{\sigma}^2}{\sigma^2} = \frac{(\underline{X} - X\underline{\hat{\beta}})\ (\underline{X} - X\underline{\hat{\beta}})'}{\sigma^2}$$

where $\hat{\beta}$ satisfies (3.53), is distributed independently of $\hat{\beta}_k$ as a chisquare with N-p degrees of freedom. Note that the assumption of estimability of each component of $\underline{\beta}$ implies that X is of rank p. Therefore

$$\frac{(N-p)^{\frac{1}{2}}(\hat{\beta}_k-\beta_k)}{\sqrt{N\,C_{kk}\,\hat{\sigma}^2}}$$

has the Student t-distribution with N-p degrees of freedom. Since the Student t-distribution is symmetric about the origin we get

$$P\{-t_{\alpha/2,N-p} \le \frac{(N-p)^{\frac{1}{2}}(\hat{\beta}_k-\beta_k)}{\sqrt{N\,C_{kk}\,\hat{\sigma}^2}} \le t_{\alpha/2,N-p}\}$$

$$= P\{\hat{\beta}_k - t_{\alpha/2,N-p}\left(\frac{NC_{kk}\,\hat{\sigma}^2}{N-p}\right)^{\frac{1}{2}} \le \beta_k \le \hat{\beta}_k + t_{\alpha/2,N-p}\left(\frac{NC_{kk}\,\hat{\sigma}^2}{N-p}\right)^{\frac{1}{2}}\}$$

$$= 1-\alpha. \tag{3.59}$$

Thus the $(1-\alpha)\%$ confidence interval of β_k is given by

$$\hat{\beta}_k - t_{\alpha/2,N-p}\left(\frac{NC_{kk}\,\hat{\sigma}^2}{N-p}\right)^{\frac{1}{2}},\quad \hat{\beta}_k + t_{\alpha/2,N-p}\left(\frac{NC_{kk}\,\hat{\sigma}^2}{N-p}\right)^{\frac{1}{2}} \tag{3.60}$$

For any estimable linear parametric function $\underline{L}'\underline{\beta}$, by Theorem 3.12,

$$\frac{(N-r)^{\frac{1}{2}}(\underline{L}'\hat{\underline{\beta}} - \underline{L}'\underline{\beta})}{\sqrt{N\hat{\sigma}^2\,\underline{L}(X'X)^{-}\underline{L}}}$$

has the Student t-distribution with N-r degrees of freedom. We are assuming tacitly that the rank of X is r. Proceeding in the same way as in (3.59) and (3.60), the $(1-\alpha)$-percent confidence interval of $\underline{L}'\underline{\beta}$ is given by

$$(\underline{L}'\hat{\underline{\beta}} - t_{\alpha/2,N-r}\left(\frac{N\hat{\sigma}^2\underline{L}'(X'X)^{-}\underline{L}}{N-r}\right)^{\frac{1}{2}},\quad \underline{L}'\hat{\underline{\beta}} + t_{\alpha/2,N-r}\left(\frac{N\hat{\sigma}^2\underline{L}'(X'X)^{-}\underline{L}}{N-r}\right)^{\frac{1}{2}})$$

In the case where the rank of X is p, N-r should be replaced by N-p and $(X,X)^{-}$ by $(X'X)^{-1}$.

Let us now consider k linearly independent estimable linear parametric functions

$$\phi_1 = \underline{L}_1'\underline{\beta},\ldots,\phi_k = \underline{L}'_k\underline{\beta} \tag{3.61}$$

which in matrix notation may be written as

$$\underline{\phi} = L'\underline{\beta}\,, \tag{3.62}$$

where $\underline{\phi} = (\phi_1,\ldots,\phi_k)'$, L' is a $k \times p$ matrix with rows $\underline{L}'_i$, $i = 1,\ldots,k$. The least square estimator $\hat{\phi}$ of ϕ is given by

$$\hat{\underline{\phi}} = (\hat{\phi}_1,\ldots,\hat{\phi}_k)' = L\hat{\underline{\beta}}$$

where $\hat{\underline{\beta}}$ is given in (3.18) and $\hat{\phi}_i = \underline{L}'_i\hat{\underline{\beta}}$. Obviously

$$E(\hat{\underline{\phi}}) = \underline{\phi}\,,$$

and from Lemma 3.9

$$\mathrm{Cov}(\hat{\underline{\phi}}) = \sigma^2(\underline{L}'_i(X'X)^-\underline{L}_j) = \sigma^2 V \text{ (say)}.$$

From Theorem 3.12, corollary 1,

$$\frac{(\hat{\underline{\phi}} - \underline{\phi})'\,V^{-1}(\hat{\underline{\phi}} - \underline{\phi})}{\sigma^2}$$

is distributed independently of $N\hat{\sigma}^2$ as χ^2_k and $N\hat{\sigma}^2/\sigma^2$ is distributed as χ^2_{N-r}. Thus

$$\frac{(N-r)(\hat{\underline{\phi}} - \underline{\phi})'\,V^{-1}(\hat{\underline{\phi}} - \underline{\phi})}{k\,N\,\hat{\sigma}^2} \tag{3.63}$$

is distributed as a central F with $(k, N-r)$ degrees of freedom. Hence

$$P\left\{\frac{(N-r)(\hat{\underline{\phi}}-\underline{\phi})'V^{-1}(\hat{\underline{\phi}}-\underline{\phi})}{k\,N\,\hat{\sigma}^2} \le F_{1-\alpha,k,N-r}\right\} = 1-\alpha.$$

The set of all $\underline{\phi}$ satisfying

$$\frac{(N-r)(\hat{\underline{\phi}}-\underline{\phi})'\,V^{-1}(\hat{\underline{\phi}}-\underline{\phi})}{k\,N\,\hat{\sigma}^2} \le F_{1-\alpha,k,N-r} \tag{3.64}$$

for given $\hat{\underline{\phi}}$, $\hat{\sigma}^2$ is an ellipsoid with centre $\hat{\phi}$ and is called the $(1-\alpha)$-percent confidence set of $\underline{\phi}$. Note that the probability that this random ellipsoid covers the true parameter point $\underline{\phi}$ is $1-\alpha$, no matter what the values of unknown parameters $\underline{\beta}, \sigma^2$ are.

3.4. TESTS OF HYPOTHESES IN THE REGRESSION MODEL

General linear hypotheses are hypotheses concerning parameters $\underline{\beta}$ of regression model (3.9) with the added restriction that $X_1, \ldots, X_N$ are independently normally distributed with means and variances as given in (3.9). The technique which is mostly used in testing these hypotheses is called the analysis of variance technique and is sometimes abbreviated as ANOVA. This involves in breaking down the total variation into orthogonal components representing variations due to different cause factors. Let $\underline{X} = (X_1, \ldots, X_N)'$ be normally distributed with $E(\underline{X}) = X\underline{\beta}$ and $\mathrm{Cov}(\underline{X}) = \sigma^2 I$. We shall consider two different categories of testing problems. Under the first category we will consider the following two:

(a) To test the null hypothesis $H_o: \underline{\beta} = \underline{\beta}^o = (\beta_1^o, \ldots, \beta_p^o)'$, where $\underline{\beta}^o$ is specified and σ^2 is unknown.

(b) To test the null hypothesis H_o: $\beta_k = \beta_k^o$, $k = 1, \ldots, q < p$ when $\beta_{q+1}, \ldots, \beta_p$ and σ^2 are unknown.

For these two problems we shall assume that all β_k are estimable. In other words X is of full rank p.

Under the second category we shall consider the following two testing problems:

(c) To test the null hypothesis $H_o: \underline{L}'\underline{\beta} = \delta$ where δ is a specified constant and $\underline{L}'\underline{\beta}$ is estimable.

(d) To test the null hypothesis H_o that k estimable linear parametric functions ϕ_i have preassigned specific values δ_i. In the notations of (3.60) and (3.61)

$$H_o: \underline{\phi} = \underline{\delta}$$

where $\underline{\delta} = (\delta_1, \ldots, \delta_k)'$.

3.4.1. Likelihood Ratio Tests: We shall use the likelihood test principle to derive the test statistics in the first two cases. In the other two cases we shall give the commonly used optimum test procedures.

Case (a). The likelihood of the observation vector $\underline{x} = (x_1, \ldots, x_N)'$ is given by

$$L(\underline{x} \mid \underline{\beta}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\{-\frac{1}{2\sigma^2}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})\}. \quad (3.65)$$

Given $\underline{x}$, L is a function $L(\underline{\beta}, \sigma^2)$ of $\underline{\beta}$ and σ^2 only. The parametric space Ω is the p + 1 dimensional space

$\Omega = \{(\underline{\beta}, \sigma^2) : -\infty < \beta_i < \infty, \ \sigma^2 > 0\}$. Since $L(\underline{\beta}, \sigma^2)$ is maximum when

$$\underline{\beta} = \hat{\underline{\beta}} = (X'X)^{-1}X'\underline{x}$$

$$\sigma^2 = \hat{\sigma}^2 = \frac{1}{N}(\underline{x} - X\hat{\underline{\beta}})' \ (\underline{x} - X\hat{\underline{\beta}}) \ ,$$

we get

$$\max_{\Omega} L(\underline{\beta}, \sigma^2) = \left(\frac{N}{2\pi(\underline{x}-X\hat{\underline{\beta}})'(\underline{x}-X\hat{\underline{\beta}})}\right)^{\frac{1}{2}N} \exp\{-\tfrac{1}{2}N\} . \quad (3.66)$$

Under the null hypothesis H_o, Ω reduces to the one dimensional space

$$\omega = \{(\underline{\beta}^o, \sigma^2) : \sigma^2 > 0\}$$

and

$$\max_{\omega} L(\underline{\beta}, \sigma^2) = \left(\frac{N}{2\pi(\underline{x}-X\underline{\beta}^o)'(\underline{x}-X\underline{\beta}^o)}\right)^{\frac{1}{2}N} \exp\{-\tfrac{1}{2}N\} . \quad (3.67)$$

Thus the likelihood ratio λ is given by

$$\lambda = \frac{\max_{\omega} L(\underline{\beta}, \sigma^2)}{\max_{\Omega} L(\underline{\beta}, \sigma^2)} = \left(\frac{(\underline{x}-X\hat{\underline{\beta}})'(\underline{x}-X\hat{\underline{\beta}})}{(\underline{x}-X\underline{\beta}^o)'(\underline{x}-X\underline{\beta}^o)}\right)^{\frac{1}{2}N}$$

$$= \left(1 + \frac{(\hat{\underline{\beta}} - \underline{\beta}^o)'X'X(\hat{\underline{\beta}} - \underline{\beta}^o)}{(\underline{x}-X\hat{\underline{\beta}})'(\underline{x}-X'\hat{\underline{\beta}})}\right)^{-\frac{1}{2}N} \quad (3.68)$$

$$= (1 + \frac{q_1}{q_2})^{-\frac{1}{2}N}$$

where

$$\begin{aligned} q_1 &= (\underline{\hat{\beta}} - \underline{\beta}^o)'\, X'X(\underline{\hat{\beta}} - \underline{\beta}^o) \\ &= (\underline{x} - X\underline{\beta}^o)'\, X(X'X)^{-1}X'(\underline{x} - X\underline{\beta}^o) \ , \\ q_2 &= (\underline{x} - X\underline{\hat{\beta}})'(\underline{x} - X\underline{\hat{\beta}}) = \underline{x}'(I - X(X'X)^{-1}X')\underline{x} \\ &= (\underline{x} - X\underline{\beta}^o)'(I - X(X'X)^{-1}X')(\underline{x} - X\underline{\beta}^o) \ . \end{aligned}$$

Since λ is a monotonically decreasing function of q_1/q_2, the likelihood ratio test rejects H_o whenever

$$\frac{q_1}{q_2} \geq C \tag{3.69}$$

where the constant C depends on the size α of the test.

The test statistic q_1/q_2 can also be obtained by the method of least squares. Let us first observe that $q_1 = (\underline{\hat{\beta}} - \underline{\beta}^o)'X'X(\underline{\hat{\beta}}-\underline{\beta}^o)$

$$= \min_{\omega}\ (\underline{x} - X\underline{\beta})'(\underline{x} - X\underline{\beta}) - \min_{\Omega}\ (\underline{x} - X\underline{\beta})'(x - X\underline{\beta}) \tag{3.70}$$

In the terminology of least squares

$$\min_{\omega}\ (\underline{x} - X\underline{\beta})'(\underline{x} - X\underline{\beta}) \text{ and } \min_{\Omega}\ (\underline{x} - X\underline{\beta})'(\underline{x} - X\underline{\beta})$$

are called the sum of squares due to the null hypothesis (or the total sum of squares) and the error sum of squares, respectively and q_1 is called the sum of squares due to the deviation from the null hypothesis or the sum of squares due to $\underline{\beta}$. For testing H_o we take the ratio of the sum of squares due to deviation from H_o to the error sum of squares as the appropriate test statistic.

Theorem 3.15. Let $\underline{Z} = \underline{X} - X\underline{\beta}^o$, $Q_1 = \underline{Z}'X(X'X)^{-1}X'\underline{Z}$ and $Q_2 = \underline{Z}'(I-X(X'X)^{-1}X')\underline{Z}$. Then Q_1/σ^2 and Q_2/σ^2 are independently distributed as chisquares with p and N-p degrees of freedom, respectively, when H_o is true.

Proof. Since under H_o, $\underline{Z}$ has an N-variate normal distribution with mean $\underline{0}$ and covariance matrix $\sigma^2 I$, and $X(X'X)^{-1}X'$, $I - X(X'X)^{-1}X'$ are idempotent matrices of rank p and N-p respectively, by Theorem 2.7, Q_1/σ^2 and Q_2/σ^2 are distributed (under H_o) as chisquares with p and (N-p) degrees of freedom, respectively. Furthermore, since

$$(I-X(X'X)^{-1}X')\ X(X'X)^{-1}X' = 0 \ ,$$

we conclude from Theorem 2.10 that under H_o, Q_1, Q_2 are independent.

Q.E.D.

Thus, under H_o, $\frac{N-p}{p} \frac{Q_1}{Q_2}$ has the F-distribution with p and N-p degrees of freedom. Hence the constant C in (3.69) is given by $C = F_{1-\alpha, p, N-p}$. The computation of the test statistic is usually carried out in the form given in Table 3.2.

Table 3.2. ANOVA for testing H_o: $\underline{\beta} = \underline{\beta}^o$

Sources of variation	Degrees of freedom	Sum of squares	F-value
Due to $\underline{\beta}$	p	q_1	$\frac{N-p}{q} \frac{q_1}{q_2}$
Error	N-p	q_2	
Total	N	$(\underline{x}-X\underline{\beta}^o)'(\underline{x}-X\beta^o)$	

Case (b). Let X be partitioned as $X = (X_1, X_2)$ where X_1 is the $N \times q$ submatrix of X containing the first q columns of X and X_2 contains the remaining p-q columns of X and let

$$\underline{\beta}_{(1)} = (\beta_1, \ldots, \beta_q)', \quad \underline{\beta}_{(2)} = (\beta_{q+1}, \ldots, \beta_p)' .$$

Then

$$X\underline{\beta} = X_1\underline{\beta}_{(1)} + X_2\underline{\beta}_{(2)} .$$

We are now interested in testing the null hypothesis

$$H_o: \underline{\beta}_{(1)} = \underline{\beta}^o_{(1)} = (\beta^o_1, \ldots, \beta^o_q)' ,$$

where $\underline{\beta}_{(2)}$, σ^2 are unknown. The parametric space Ω is the same as that of case (a). Under H_o, it reduces to

$$\omega = \{(\beta^o_{(1)}, \beta_{(2)}, \sigma^2): -\infty < \beta_i < \infty,\ i = q+1, \ldots, p,\ \sigma^2 > 0\}.$$

The maximum of the likelihood function $L(\underline{\beta}, \sigma^2)$ of (3.65) under Ω is given in (3.66). To find the maximum of $L(\underline{\beta}, \sigma^2)$ under ω we proceed as follows:

$$(\underline{x} - \underline{\beta})'(\underline{x} - X\underline{\beta})$$

$$= (\underline{x} - X_1\underline{\beta}_{(1)} - X_{(2)}\underline{\beta}_{(2)})'(\underline{x} - X_1\beta_{(1)} - X_2\beta_{(2)}) .$$

Thus, under H_o, the likelihood of $\underline{x}$ is given by

$$(2\pi\,\sigma^2)^{-\frac{1}{2}N}\ \{\exp -\frac{1}{2\sigma^2}\,(\underline{u}-X_2\underline{\beta}_{(2)})'(\underline{u}-X_2\underline{\beta}_{(2)})\} \qquad (3.71)$$

where $\underline{u} = \underline{x} - X_1\underline{\beta}^o_{(1)}$ and $\underline{\beta}_{(2)}, \sigma^2$ are unknown parameters. Comparing (3.71) with (3.65), it is easy to conclude that the maximum likelihood estimates of $\underline{\beta}_{(2)}$ and σ^2 under H_o are

$$\hat{\underline{\beta}}_{(2)} = (X_2'X_2)^{-1}\,X_2'\underline{u} \qquad (3.72)$$

$$\hat{\sigma}^2 = \frac{1}{N}\,(\underline{u} - X_2\hat{\underline{\beta}}_2)'\ (\underline{u} - X_2\hat{\underline{\beta}}_{(2)}\ ) .$$

It may be noted that $X_2'X_2$ being a principal minor of the positive definite matrix $X'X$, is positive definite and hence has a unique inverse $(X_2'X_2)^{-1}$. Thus from (3.71) and (3.72)

$$\max_{\omega} L(\hat{\underline{\beta}},\sigma^2) = \left(\frac{N}{2\pi\,(\underline{u}-X_2\hat{\underline{\beta}}_{(2)})'\ (\underline{u} - X_2\hat{\underline{\beta}}_{(2)})}\right)^{\frac{1}{2}N} \exp\{-\tfrac{1}{2}N\} \qquad (3.73)$$

Now

$$(\underline{u} - X_2\hat{\underline{\beta}}_{(2)})'(\underline{u}-X_2\hat{\underline{\beta}}_{(2)})' = \underline{u}'(I - X_2(X_2'X_2)^{-1}X_2')\,\underline{u}$$

$$=(\underline{x}-X_1\underline{\beta}^o_{(1)})\,(I-X_2(X_2'X_2)^{-1}X_2')\,(\underline{x}-X_1\underline{\beta}^o_{(1)})$$

$$=(\underline{x}-X_1\underline{\beta}^o_{(1)}-X_2\underline{\beta}_{(2)})'(I-X_2(X_2'X_2)^{-1}X_2')\,(\underline{x}-X_1\underline{\beta}^o_{(1)}-X_2\underline{\beta}_{(2)}) .$$

From (3.66) and (3.73), the likelihood ratio λ for testing H_o: $\underline{\beta}_{(1)} = \underline{\beta}^o_{(1)}$ is given by

$$= \frac{\max_{\omega} L(\underline{\beta},\sigma^2)}{\max_{\Omega} L(\underline{\beta},\sigma^2)} = \left(\frac{(\underline{x} - X\hat{\underline{\beta}})'\,(\underline{x} - X\hat{\underline{\beta}})}{(\underline{u} - X_2\hat{\underline{\beta}}_{(2)})'\,(\underline{u} - X_2\hat{\underline{\beta}}_{(2)})}\right)^{\frac{1}{2}N}$$

$$= (1 + \frac{q_1}{q_2})^{-\frac{1}{2}N} \qquad (3.74)$$

where

$$q_2 = (\underline{x} - X\hat{\underline{\beta}})'\ (\underline{x} - X\hat{\underline{\beta}})$$

$$= \underline{x}'(I-X(X'X)^{-1}X')\,\underline{x}$$

$$= (\underline{x}-X_1\underline{\beta}^o_{(1)}-X_2\underline{\beta}_{(2)})'(I-X(X'X)^{-1}X')\,(\underline{x}-X_1\underline{\beta}^o_{(1)}-X_2\underline{\beta}_{(2)}) ,$$

$$\begin{aligned} q_1 &= (\underline{u}-X_2\hat{\underline{\beta}}_{(2)})'(\underline{u}-X_2\hat{\underline{\beta}}_{(2)}) - (\underline{x}-X\hat{\underline{\beta}})'(\underline{x}-X\hat{\underline{\beta}}) \\ &= \underline{u}'(I-X_2(X_2'X_2)^{-1}X_2)\underline{u} - \underline{x}'(I-X(X'X)^{-1}X')\underline{x} \\ &= (\underline{x}-X_1\underline{\beta}^{o}_{(1)}-X_2\underline{\beta}_{(2)})'(X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2) \\ &\quad (\underline{x}-X_1\underline{\beta}^{o}_{(1)}-X_2\underline{\beta}_{(2)}) \ . \end{aligned} \tag{3.75}$$

The computation in (3.75) follows from the fact that

$$X'(I-X(X'X)^{-1}X') = 0$$

implies

$$X_1'(I-X(X'X)^{-1}X') = 0, \quad X_2'(I-X(X'X)^{-1}X') = 0 \ . \tag{3.76}$$

Since λ is a monotonically decreasing function of q_1/q_2, the likelihood ratio test rejects H_o whenever

$$\frac{q_1}{q_2} > C \ , \tag{3.77}$$

where the constant C depends on the size α of the test.

As in case (a) the likelihood ratio test statistic q_1/q_2 in (3.77) can also be obtained by the method of least squares. Obviously $q_1 + q_2$ is the minimum value of $(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})$ when the null hypothesis H_o: $\underline{\beta}_{(1)} = \underline{\beta}^{o}_{(1)}$ holds good and hence it represents the sum of squares due to H_o. Since q_2 represents the error sum of squares, in our present notation q_1 will represent the sum of squares due to the deviation from the null hypothesis H_o or the sum of squares due to $\underline{\beta}_{(1)}$ adjusted for $\underline{\beta}_{(2)}$. If $\underline{\beta}^{o}_{(1)} = \underline{0}$ then

$$\begin{aligned} q_1 &= \underline{x}'\underline{x} - \hat{\underline{\beta}}'_{(2)} X_2'\underline{x} - \underline{x}'\underline{x} + \hat{\underline{\beta}}'X'\underline{x} \\ &= \hat{\underline{\beta}}'X'\underline{x} - \hat{\underline{\beta}}'_{(2)} X_2'\underline{x} \ . \end{aligned} \tag{3.78}$$

The quantity $\hat{\underline{\beta}}'X'\underline{x}$ will be called the reduction sum of squares due to $\underline{\beta}$ or simply the sum of squares due to $\underline{\beta}$. Similarly $\hat{\underline{\beta}}'_{(2)} X_2'\underline{x}$ will be called the sum of squares due to $\underline{\beta}_{(2)}$ ignoring $\underline{\beta}_{(1)}$.

Theorem 3.16. Let $\underline{Z} = \underline{X} - X_1\underline{\beta}^o_{(1)} - X_2\underline{\beta}_{(2)}$, $Q_2 = \underline{Z}'A\underline{Z}$ and $Q_1 = \underline{Z}'B\underline{Z}$ where $A = I-X(X'X)^{-1}X'$. $B = X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2'$.

Then Q_1/σ^2, Q_2/σ^2 are independently distributed as chisquares with q and N-p degrees of freedom, respectively.

Proof. Since under H_o, $\underline{Z}$ has an N-variate normal distribution with mean $\underline{0}$ and covariance matrix $\sigma^2 I$, and B, A are idempotent matrices of rank q, N-p respectively, by Theorem 2.7 Q_1/σ^2 and Q_2/σ^2 are distributed under H_o as chisquares with q and N-p degrees of freedom respectively. Furthermore, since BA=0, we conclude from Theorem 2.10 that under H_o, Q_1, Q_2 are independent.

Q.E.D.

Thus, under H_o, $\frac{N-p}{q}\,\frac{Q_1}{Q_2}$ has an F-distribution with q and N-p degrees of freedom. Hence the constant C in (3.77) is given by $C = F_{1-\alpha,q,N-p}$. For practical applications $\underline{\beta}^o_{(1)} = 0$ is quite useful. We shall give below the ANOVA table for H_o: $\underline{\beta}_{(1)} = \underline{0}$.

Table 3.3. ANOVA for testing H_o: $\underline{\beta}_{(1)} = \underline{0}$

Sources of variation	Degrees of freedom	Sum of squares	F-values
Due to $\underline{\beta}$	p	$\hat{\underline{\beta}}'X'\underline{x}$	
Due to $\underline{\beta}_{(2)}$ (unadjusted)	p-q	$\hat{\underline{\beta}}_{(2)}X_2'\underline{x}$	
Due to $\underline{\beta}_{(1)}$ (adjusted)	q	q_1	$\frac{N-p}{q}\,\frac{q_1}{q_2}$
Error	N-p	q_2	
Total	N	$\underline{x}'\underline{x}$	

Case (c). For any estimable linear parametric function $\underline{L}'\underline{\beta}$, by Theorem (3.14), under $H_o: \underline{L}'\underline{\beta} = \delta$,

$$t = \frac{(n-r)\,(\underline{L}'\hat{\underline{\beta}} - \delta)}{(N\hat{\sigma}^2\,\underline{L}'(X'X)^-\underline{L})^{\frac{1}{2}}} \tag{3.79}$$

has the Student t-distribution with N-r degrees of freedom. Hence an appropriate test for testing H_o against alternatives $\underline{L}'\underline{\beta} \neq \delta$ rejects H_o whenever

$$|t| \geq t_{1-\alpha/2,N-p} \; . \tag{3.80}$$

Case (d). For testing $H_o:\underline{\phi} = \underline{\delta}$, an appropriate test procedure is to reject H_o whenever (see (3.63))

$$\frac{(N-r)\,(\hat{\underline{\phi}}-\underline{\delta})-V^{-1}(\hat{\underline{\phi}}-\underline{\delta})}{k\,N\,\hat{\sigma}^2} \geq F_{1-\alpha,k,N-r} \tag{3.81}$$

3.5. ONE-WAY CLASSIFICATIONS WITH LINEAR MODELS OF FULL RANK

It refers to the comparison of means of several univariate normal populations with the same unknown variance σ^2. Let X_{ij}, $j = 1,\ldots,N_i$ be a random sample of size N_i from the ith normal population with mean β_i and variance σ^2, $i = 1,2,\ldots,p$, and let the random samples from different populations be independent of each other. Write

$$N = \sum_{i=1}^{p} N_i \, ,$$

$$\underline{X} = (X_{11},\ldots,X_{1N_1},\ldots, X_{p1},\ldots,X_{pN_p})' \, ,$$

$$\underline{x} = (x_{11},\ldots,x_{1N_1},\ldots,x_{p1},\ldots,x_{pN_p})' \, ,$$

$$\underline{\beta} = (\beta_1,\ldots,\beta_p)' \, .$$

The N-vector $\underline{X}$ has a multivariate normal distribution with

$$E(\underline{X}) = X\underline{\beta} \, , \; Cov(\underline{X}) = \sigma^2 I, \tag{3.82}$$

where the $N \times p$ matrix X has the following form: the first N_1 rows are all $\varepsilon_1' = (1,0,\ldots,0)$, the next N_2 rows are all $\varepsilon_2' = (0,1,0,\ldots,0),\ldots,$ and the last N_p rows are all $\varepsilon_p' = (0,\ldots,0,1)$. Obviously X has rank p and hence (3.82) represents a fixed effect linear model of full rank.

We are interested here in testing the null hypothesis H_o: $\beta_1 = \ldots = \beta_p = \beta(\text{say})$, where β, σ^2 are unknown.

The parametric space Ω is the (p+1) dimensional space given by

$$\Omega = \{(\underline{\beta},\sigma^2) : -\infty < \beta_i < \infty,\ i = 1,\dots,p;\ \sigma^2 > 0\}.$$

Under H_o it reduces to the 2-dimensional space

$$\omega = \{(\beta,\sigma^2) : -\infty < \beta < \infty, \sigma^2 > 0\} .$$

Let $E(X_{ij}) = \mu_{ij}$. With the notation of Example 3.2 the hypothesis H_o is linear hypothesis with s-r = 1, (s=p), with π_Ω given by $\mu_{ij} = \beta_i$, $j = 1,\dots,N_i$, $i = 1,\dots,p$, and with π_ω the line on which all the N coordinates of $\underline{\mu} = (\mu_{11},\dots,\mu_{1N_1},\dots,\mu_{p1},\dots,\mu_{pN_p})'$ are equal.

Let $x_{..} = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{N_i} x_{ij}$, $x_{i.} = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}$. Since

$$\sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-\beta_i)^2 = \sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-x_{i.})^2 + \sum_{i=1}^{p} N_i(x_{i.}-x_{..})^2 ,$$

we get the least squares estimate $\hat{\mu}_{ij} = x_{i.}$ of μ_i, when π_Ω holds.

Also

$$\sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-\beta)^2 = \sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-x_{..})^2 + N(x_{..}-\beta)^2 .$$

The least square estimate $\hat{\hat{\mu}}_{ij}$ or μ_{ij} under π_ω is given by $\hat{\hat{\mu}}_{ij} = x_{..}$.

Thus from (3.7), the test statistic w is given by

$$w = \frac{N-p}{p-1}\,\frac{\sum_{j=1}^{p} N_i(x_{i.}-x_{..})^2}{\sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-x_{i.})^2} \tag{3.83}$$

and the distribution of w depends only on the parameter $\delta^2 =$ $\sum_{i=1}^{p} N_i \frac{(\beta_i-\beta_.)^2}{\sigma^2}$, where $\beta_. = \sum_{i=1}^{p} N_i\beta_i / N$. Thus the uniformly most powerful invariant test of H_o, against the alternatives $\delta^2 > 0$, rejects H_o whenever $w \geq c$, where the constant c depends on the size α of the tests. We shall show later that $c = F_{1-\alpha, p-1, N-p}$.

We now treat the problem as the regression model and find the likelihood ratio rest statistic for this problem.

The likelihood function is given by

$$L(\underline{\beta}, \sigma^2) = (2\pi)^{-\frac{1}{2}N} (\sigma^2)^{-\frac{1}{2}N} \exp \{- \frac{1}{2\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - \beta_i)^2\} .$$

It can be easily verified that

$$\max_{\omega} L(\underline{\beta}, \sigma^2) = \left(\frac{N}{2\pi \left(\sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - x_{..})^2 \right.} \right)^{\frac{1}{2}N} \exp \{-\tfrac{1}{2}N\} ,$$

$$\max_{\hat{\Omega}} L(\underline{\beta}, \sigma^2) = \left(\frac{N}{2\pi \left(\sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - x_{i.})^2 \right.} \right)^{\frac{1}{2}N} \exp \{-\tfrac{1}{2}N\} .$$

Since

$$\sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - x_{..})^2 = \sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - x_{i.})^2 + \sum_{i=1}^{p} N_i (x_{i.} - x_{..})^2 ,$$

the likelihood ratio λ for testing H_o can be written as

$$\lambda = \frac{\max_{\omega} L(\underline{\beta}, \sigma^2)}{\max_{\Omega} L(\underline{\beta}, \sigma^2)} = \left(1 + \frac{q_1}{q_2}\right)^{-\frac{1}{2}N} , \tag{3.84}$$

where

$$q_1 = \sum_{i=1}^{p} N_i (x_{i.} - x_{..})^2, \quad q_2 = \sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij} - x_{i.})^2 .$$

In the terminology of least squares, q_1 represents the sum of squares due to the deviation from the null hypothesis H_o or the between populations sum of squares and q_2 represents the error Sum ~~sum~~ of squares (within populations sum of squares). The quantity ~~quantity~~ $q_1 + q_2$ represents the sum of squares due to the null hypothesis H_o or the total sum of squares. From (3.84) the likelihood ratio test rejects H_o whenever

$$\frac{N-p}{p-1}\frac{q_1}{q_2} \geq c , \tag{3.85}$$

where the constant c depends on the size α of the test. This test has also been obtained in (3.83) by using invariance consideration. This test can also be derived from the least squares theory by taking the ratio of $[\min_{\omega}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})-\min_{\Omega}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})]$ to $\min_{\Omega}(\underline{x}-X\underline{\beta})'(\underline{x}-X\underline{\beta})$.

Theorem 3.17. Let $Q_1 = \sum_{i=1}^{p} N_i(X_{i.}-X_{..})^2$, $Q_2 = \sum_{i=1}^{p} S_i^2$, $S_i^2 = \sum_{j=1}^{N_i}(X_{ij}-X_{i.})^2$, $X_{..} = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{N_i} X_{ij}$ and $X_{i.} = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}$. Then ~~then~~ Q_1/σ^2, Q_2/σ^2 are independently distributed as chisquare with p-1, N-p degrees of freedom, respectively, when H_o is true.

Proof. Obviously for each i, $X_{i.}$, S_i^2 are independent, S_i^2/σ^2 is distributed as chisquare with N_i-1 degrees of freedom and $\sqrt{N_i}X_{i.}$ is normally distributed with mean $\sqrt{N_i}\beta$ and variance σ^2 when H_o is true. Since random samples from different populations are independent of each other we conclude that Q_2/σ^2 is distributed as chisquare with N-p degrees of freedom and is independent of Q_1. Furthermore under H_o it is easy to see that Q_1/σ^2 is distributed as chisquare with p-1 degrees of freedom.

Q.E.D.

Thus, under H_o, $\frac{N-p}{p-1}\frac{Q_1}{Q_2}$ has the F-distribution with p-1,N-p degrees ~~egrdes~~ of freedom. Hence the constant c in (3.83) and (3.85) is equal to $F_{1-\alpha,p-1,N-1}$.

Table 3.4. ANOVA for One-way classification

Sources of variation	Degrees of freedom	Sum of squares	Expected mean sum of squares	F
Between populations	p-1	q_1	$\frac{1}{p-1}\sum_{i=1}^{p}(\beta_i-\bar{\beta})^2$	$\frac{N-p}{p-1}\frac{q_1}{q_2}$
Within populations	N-p	q_2	σ^2	
Total	N-1	$\sum_{i=1}^{p}\sum_{j=1}^{N_i}(x_{ij}-x_{..})^2$		

If the F-value is greater than $F_{1-\alpha,p-1,N-p}$, we say that it is significant at level α and consequently we reject the null hypothesis $H_o:\beta_1=\ldots=\beta_p$. Naturally such rejection leads to further investigation to decide which means are different. Since

$$\frac{x_{i.}-x_{k.}}{\sqrt{s^2\left(\frac{1}{N_i}+\frac{1}{N_k}\right)}},$$

where $s^2 = Q_2/(N-p)$, has the Student t-distribution with N-p degrees of freedom, the null hypothesis that $\beta_i = \beta_k$ will be rejected at level α if the observed difference $x_{i.} - x_{k.}$ is greater than the critical difference

$$t_{1-\alpha/2,N-p}\sqrt{s^2\left(\frac{1}{N_i}+\frac{1}{N_k}\right)} \tag{3.86}$$

where $s^2 = q_2/(N-p)$. Thus we may compute all possible critical differences arising out of all possible pairs (β_i,β_k), $i \neq k = 1,\ldots,p$, and compare them with their corresponding observed differences. On the basis of such comparisons it will be possible to divide the p populations into different groups such that the populations in the same group have the same mean and those in different groups have different means. The computation is simplified if all N are equal to n (say). In subh a case we compute the common critical difference

$$ccd = t_{1-\alpha/2,N-p}\sqrt{\frac{2s^2}{n}} \tag{3.87}$$

and compare it with the observed differences $x_{i.}-x_{k.}$, for all $i \neq k$. If any difference exceeds the ccd, then the corresponding means are said to come from populations with different means.

When comparing a number of population means one would thus typically wish to decide not only whether the means are equal, but in case the hypothesis of equality of means is rejected, would like to rank or divide them into subgroups such that the populations with the same means belong to the same subgroup and with different means to different subgroups. This we have achieved by the combined use of F and t-tests. Other types of procedures for this have been proposed by, among others, Student-Newman-Keuls, the "range test", Duncan (1955) "the multiple range and F test"; Tukey (1951, 1953), "a test based on allowances" and Scheffe (1953), "S and T methods for judging all contrasts in the analysis of variance". These procedures are commonly called multiple Comparison ~~comparison~~ procedures. We shall discuss only Student-Newman-Keuls test and Duncan's multiple range test. For a detailed discussion on multiple comparison procedures and their relative merits and demerits the reader is referred to Federer (1955). Though we will discuss these two procedures for comparison of means in the one-way classification only, it can be applied to similar other problems without any difficulty. Let the observed means $x_{1.},\ldots,x_{p.}$ be ranked from the highest to the smallest as $x_p,\ldots,x_1$, i.e., x_1 = the smallest of $x_{1.},\ldots,x_{p.}$; ...; x_i = the ith largest of $x_{1.},\ldots,x_{p.}$; x_p = the largest of $x_{1.},\ldots,x_{p.}$. The quantity $x_p - x_1$ is called the range and

$$q_{p,N-p} = \frac{(x_p - x_1)\sqrt{n}}{s} \tag{3.88}$$

where $N_1 = \ldots = N_p = n$ (by assumption) is called the Studentized range of the observed means $x_{1.},\dots,x_{p.}$. Let $q_{\alpha,p,v}$ denote the upper α-percent point of the Studentized range (here $v = N-p$). The values of $q_{\alpha,p,v}$ are given in Table 3.6.

3.5.1. Student-Newman-Keuls Test (Student (1927), Newman (1939), Keuls (1952)): We compute the upper α-percent point critical Studentized range W_p given by (as a function of p for given α, v)

$$W_p = q_{\alpha,p,v}\,\frac{s^2}{n} \, , \qquad (3.89)$$

Now we compare the observed range $x_p - x_1$ with W_p. If it is less than W_p, the process of comparison stops and we conclude that the means $\beta_1,\dots,\beta_p$ are equal. If $x_p - x_1$ exceeds W_p, we divide the ranked means into two subgroups of p-1 ranked means each, x_p to x_2 and x_{p-1} to x_1. Then we compare the ranges $x_p - x_2$, $x_{p-1} - x_1$ with W_{p-1}. If either range is less than W_{p-1}, the means in the corresponding group are equal. If either range exceeds W_{p-1}, then the p-1 means in the group concerned are again divided into two groups of p-2 means each and the ranges are compared with W_{p-2}. This is continued until a group of i means is found whose range does not exceed W_i. In other words, by this method the ~~difference~~ differences between any two means is significant provided the range of the observed means of each and every subgroup containing the given ~~two~~ two means is significant according to Studentized critical range.

3.5.2. Duncan's Multiple Range Test: This test procedure is essentially the same as that of Student-Newman-Keuls except that here the observed ranges are compared with Duncan's α-percent critical range

$$D_p = q^*_{\alpha,p,v}\sqrt{\frac{s^2}{n}} \, , \qquad (3.90)$$

where $q^*_{\alpha,p,v}$ denotes the upper α-percent point of the Studentized range based on Duncan's range in Table 3.5, instead of W_p. ~~Duncan~~ Duncan' feels that this test is better than any previously proposed test for comparing differences between any two ranked means.

Table 3.5

Values of $q^*_{\alpha,p,v}$

$\alpha = .05$

v P	2	3	4	5	6	8	10	14	20	50	100
10	3.15	3.29	3.37	3.43	3.46	3.47	3.47	3.47	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.44	3.46	3.46	3.46	3.48	3.48	3.48
14	3.03	3.18	3.27	3.33	3.37	3.41	3.44	3.36	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.39	3.43	3.45	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.37	3.41	3.45	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.36	3.40	3.44	3.47	3.47	3.47
24	2.92	3.07	3.15	3.22	3.28	3.34	3.38	3.44	3.47	3.47	3.47
30	2.85	3.04	3.12	3.20	3.25	3.32	3.37	3.43	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.28	3.33	3.40	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.26	3.32	3.40	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.23	3.29	3.38	3.47	3.61	3.67
			$\alpha = .01$			$\alpha = .0$					
10	4.48	4.73	4.88	4.96	5.06	5.20	5.28	5.42	5.55	5.55	5.55
12	4.32	4.55	4.68	4.76	4.84	4.96	5.07	5.17	5.26	5.26	5.26
14	4.21	4.42	4.55	4.63	4.70	4.83	4.91	5.00	5.07	5.07	5.07
16	4.13	4.34	4.45	4.54	4.60	4.72	4.79	4.88	4.94	4.94	4.94
18	4.07	4.27	4.38	4.46	4.53	4.64	4.71	4.79	4.85	4.85	4.85
20	4.02	4.22	4.33	4.40	4.47	4.58	4.65	4.73	4.79	4.79	4.79
24	3.96	4.14	4.24	4.33	4.39	4.49	4.57	4.64	4.72	4.74	4.74
30	3.89	4.06	4.16	4.25	4.32	4.41	4.48	4.58	4.65	4.72	4.72
60	3.76	3.92	4.03	4.12	4.17	4.27	4.34	4.44	4.53	4.66	4.66
100	3.71	3.86	3.98	4.06	4.11	4.21	4.29	4.38	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.14	4.20	4.31	4.41	4.60	4.68

Reproduced with the permission of the editor of Biometrics from the paper by D.B. Duncan, Biometrics 11:1-42, 1955.

Table 3.6

Values of $q_{\alpha,p,v}$

$\alpha = .05$

v^p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.85	5.93	6.05	6.12	6.20	6.27	6.34	6.41	6.47
11	3.11	3.82	4.26	4.58	5.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.14	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95	6 02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.46	4.69	5.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.06	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79	5.86	5.92	5.98	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.79	5.85	5.91	5.96
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	2.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50	5.56	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.55	5.59
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.43
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17	5.22	5.27	5.32	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.23
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.84	4.89	4.93	4.97	5.01

$\alpha = .01$

v \ P	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	4.48	5.26	5.77	6.14	6.43	6.67	6.88	7.06	7.22	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	4.39	5.14	5.62	5.98	6.25	6.47	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.60	7.67	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	7.90	7.01	7.10	7.19	7.27	7.35	7.42	7.49	7.55
14	4.21	4.89	5.32	5.64	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.34	7.40
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.85	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.78	5.19	5.49	5.72	5.91	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.73	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.95	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	4.05	4.66	5.05	5.34	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.63	5.02	5.30	5.51	5.69	5.84	5.97	6.09	6.19	6.28	6.37	6.45	6.52	6.59	6.66	6.71	6.77	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.57	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.53	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.36	4.70	4.95	5.11	5.26	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.78	5.83	5.88	5.93	5.98	6.01
120	3.70	4.20	4.50	4.71	4.87	5.00	5.12	5.21	5.30	5.38	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.53	5.57	5.61	5.64

Reproduced with the permission of the Biometrika Trustees from the paper by J.M. May, "Extended and corrected tables of the upper percentage points of the 'studentized' range", Biometrika, 39: 192-193, 1952. An improved version of this table may be found in Table 29 of Biometrika tables for statisticians, Vol.1, Cambridge University Press, 1954, by E.S. Pearson and H.O. Hartley.

Example 3.5. Consider the data in Table 3.1. The analysis of variance table for this is given in Table 3.7.

Table 3.7. ANOVA for the data in Table 3.1

Sources of variation	Degrees of freedom	Sum of squares	F
Between varieties	3	18.6757	1.0834
Within varieties	52	298.7843	
Total	55	317.4600	

Since

$$F_{1-\alpha,3,5,2} = \begin{cases} 4.22 \quad, & \text{if} \quad \alpha = 0.01 \quad, \\ 3.40 \quad, & \text{if} \quad \alpha = \ .025 \ , \\ 2.80 \quad, & \text{if} \quad \alpha = \ .015 \ , \\ 2.205 \ , & \text{if} \quad \alpha = \ .10 \quad, \\ 1.415 \ , & \text{if} \quad \alpha = \ .25 \quad, \end{cases}$$

we accept the hypothesis of equality of means of different varieties at size $\alpha \le .25$.

3.5.3. Tests of Equality of Variances: Let X_{ij}, $j = 1,\ldots,N_i$, $i = 1,\ldots,p$ be a random sample size N_i from the ith population with mean β_i and variance σ_i^2 and let

$$H_o;\ \sigma_1^2 = \ldots = \sigma_p^2 \tag{3.91}$$

One can obtain the likelihood ratio test or the Bartlett modification thereof for this problem. We shall treat here an alternative approach by which it can be reduced in large samples to one-way classification with known variance.

Unfortunately for this problem neither there exists a uniformly unbiased test nor there is available a group of transformations with respect to which there exists a uniformly most powerful invariant test (see Lehmann(1959)). For large $N(= \sum_{i=1}^{p} N_i)$, however, this problem can be reduced to one-way classification. Let

$$X_{i.} = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}, \quad S_i^2 = \sum_{j=1}^{N_i} (X_{ij} - X_{i.})^2, \quad i = 1,\dots,p.$$

It is well known that $(X_{i.}, S_i^2, i = 1,\dots,p)$ is a sufficient statistic for this problem. The original hypothesis $H_0: \sigma_1^2 = \dots = \sigma_p^2$ remains invariant under the group G of transformations

$$X_{ij} \to X_{ij} + C_i, \quad j = 1,\dots,N_i, \quad i = 1,\dots,p$$

where the C_i are reals and which in the space of the sufficient statistic induces the transformations

$$X_{i.} \to X_{i.} + C_i, \quad S_i^2 \to S_i^2, \quad i = 1,\dots,p.$$

A set of maximal invariants under G in the space of the sufficient statistic is $(S_1^2,\dots,S_p^2)$.

Since S_i^2 is distributed as the sum of squares of N_i-1 independent normal random variables with means 0 and with the same variance σ_i^2, we can conclude from the central limit theorem that as $N_i \to \infty$

$$(N_i-1)^{\frac{1}{2}} \left(\frac{S_i^2}{N_i-1} - \sigma_i^2 \right)$$

is approximately distributed as a normal random variable with mean 0 and variance $2\sigma_i^4$.

This result is not convenient for our purpose as both the mean and the variance of this limiting distribution depend on σ_i^2. However, this difficulty can be avoided by considering independent random variables $Z_i = \log (S_i^2/(N_i-1))$, $i=1,\dots,p$ instead of S_i^2. For large N_i, the random variable Z_i is approximately normally distributed with mean $\theta_i = \log \sigma_i^2$ and variance $a_i^2 = 2(N_i-1)^{-1}$.

In terms of the Z_i, the problem of testing H_0 reduces to the problem of testing the equality of means of p independent normal random variables $Z_i,\dots,Z_p$, with known variances.

In the special case where all N_i are equal, the Z_i have the same known variance a^2 (say) but different means θ_i. The problem of testing the hypothesis $\theta_i = \dots \theta_p$ remains invariant under the group of transformations

$$Z_i \to Z_i + C, \quad i = 1,\dots,p,$$

where C is real and under the group of orthogonal transformations of hyperplanes which are orthogonal to $Z_1 = \dots = Z_p$. The uniformly most powerful invariant test under these groups rejects H_o whenever

$$\sum_{i=1}^{p} \frac{(z_i - \bar{z})^2}{a^2} \geq k \tag{3.92}$$

where $\bar{z} = \frac{1}{p} \sum_{i=1}^{p} z_i$, and the constant k depends on the size α of the test. Obviously under the hypothesis $\theta_1 = \dots = \theta_p$, $\sum_{i=1}^{p} (Z_i - \bar{Z})^2/a^2$ is distributed as chisquare with p-1 degrees of freedom. So $k = \chi^2_{1-\alpha,p-1}$.

In the general case the problem reduces to a general linear hypothesis with known variance by considering $Y_i = Z_i/a_i$, $i = 1,\dots,p$. The uniformly most powerful invariant test under the group of transformations leaving the problem invariant, rejects $H_o: \theta_1 = \dots = \theta_p$ whenever

$$u = \sum_{i=1}^{p} \frac{1}{a_i^2} \left(z_i - \frac{\sum_{i=1}^{p} \frac{z_i}{a_i^2}}{\sum_{i=1}^{p} (a_i^2)^{-1}} \right)^2 \geq k,$$

where the constant k depends on the size α of the tests. Since under H_o, U is distributed as χ^2_{p-1}, we get $k = \chi^2_{1-\alpha,p-1}$. It may be noted that we are only considering here the alternatives

$$\delta^2 = \sum_{i=1}^{p} \frac{1}{a_i^2}\left(\theta_i - \frac{\sum_{i=1}^{p} \frac{\theta_i}{a_i^2}}{\sum_{i=1}^{p} (a_i^2)^{-1}}\right)^2$$

$$> 0$$

for the invariant tests. Under H_o, δ^2 is equal to zero.

3.6. EXPERIMENTAL DESIGN MODELS

In this section we shall consider linear models (3.9) where the design matrix X does not have full rank and the matrix X consists of 0's and 1's only.

Let us consider once again the one-way classification. Although the linear model adopted earlier to analyse one-way classification is the most natural one for such experiments, the following alternative model resembles more closely the models used in experimental designs and at the same time it gives us an insight into the problem of design models and their analysis. Let

$$E(X_{ij}) = \mu + \alpha_i, \tag{3.94}$$

$$V(X_{ij}) = \sigma^2, \quad j = 1,\ldots,N_i, \quad i = 1,\ldots,p,$$

where X_{ij} are independently normally distributed. Clearly (3.94) can be obtained from (3.9) by writing $\mu = \bar{\beta} = \frac{1}{p}\sum_1^p \beta_i, \alpha_i = \beta_i - \bar{\beta}$. In matrix notation (3.94) can be written as

$$E(\underline{X}) = X^* \ \underline{\beta}^*, \quad \mathrm{Cov}(\underline{X}) = \sigma^2 I$$

where $\underline{\beta}^* = (\mu, \alpha_1,\ldots, \alpha_p)'$ and the design matrix X^* of dimension $N \times (p+1)$ can be obtained from the design matrix of the model (3.9) by adjoining a column of 1's on the left of X. Obviously the first column of X^* is the vector sum of all its remaining p columns and hence the rank of X^* is p. It is thus apparent that all linear parametric functions of $\alpha_1,\ldots,\alpha_p$ are not estimable. What kind of linear parametric functions are then estimable ? For any linear estimator

$$\sum_{i=1}^{p}\sum_{j=1}^{N_i} a_{ij}X_{ij}\,, \qquad \text{with} \quad c_i = \sum_{j=1}^{N_i} a_{ij}\,,$$

we get

$$E\Big(\sum_{i=1}^{p}\sum_{j=1}^{N_i} a_{ij}\, X_{ij}\Big) = \sum_{i=1}^{p}\sum_{j=1}^{N_i} a_{ij}(\mu + \alpha_i)$$

$$= \sum_{i=1}^{p} c_i\,\alpha_i + \mu\Big(\sum_{i=1}^{p} c_i\Big) .$$

Thus estimable linear parametric functions of $\underline{\beta}^*$ are of the form $\sum_{i=1}^{p} c_i\alpha_i + \mu\big(\sum_{i=1}^{p} c_i\big)$.

Furthermore we may also conclude that any linear parametric function $\sum_{i=1}^{p} c_i\alpha_i$ is estimable if and only if $\sum_{i=1}^{p} c_i = 0$, i.e. $\sum_{i=1}^{p} c_i\alpha_i$ is a contrast. Thus neither $\sum_{i=1}^{p} \alpha_i$ nor any of the parameters $\mu, \alpha_1, \ldots, \alpha_p$ is estimable.

The least squares estimates $\bar\mu$, $\bar\alpha_1, \ldots, \bar\alpha_p$ of $\mu, \alpha_1, \ldots, \alpha_p$ are given by the normal equations

$$N\bar\mu + \sum_{i=1}^{p} N_i\bar\alpha_i = \sum_{i=1}^{p} N_i x_{i.}\,, \tag{3.95}$$

$$\bar\mu + \bar\alpha_i = x_{i.}\,, \quad i = 1, \ldots, p .$$

Obviously the first equation in (3.95) can be obtained from the second one. Hence they are linearly dependent and consequently $\bar\mu, \bar\alpha_1, \ldots, \bar\alpha_p$ do not have a unique solution (the same state of affair exists for the maximum likelihood estimates). However if we impose the side condition that $\sum_{1}^{p} \bar\alpha_i = 0$, then the normal equations (3.91) have a unique solution and they are $\bar\mu = x_{..}$, $\bar\alpha_i = x_{i.} - x_{..}$. But we can without any loss of generality assume that $\sum_{1}^{p} \alpha_i = 0$. For,

$$\mu_{ij} = E(X_{ij}) = \mu + \alpha_i = (\mu + \alpha_.) + (\alpha_i - \alpha_.)$$

$= \mu^* + \alpha_i^*$

where $\alpha_{\cdot} = \frac{1}{p}\sum_1^p \alpha_i$ and $\sum_1^p \alpha_i^* = 0$. Since

$$\mu^* + \alpha_i^* = \mu_{i1} \quad \text{and} \quad \sum_1^p \alpha_i^* = 0 ,$$

$\mu^* = \mu_{\cdot 1}$, $\alpha_i^* = \mu_{i1} - \mu_{\cdot 1}$ are uniquely determined by μ_{ij}. Obviously such a unique representation of μ, α_i's in terms of μ_{ij} is not possible.

Finally, if any linear parametric function $a_0\mu^* + \sum_1^p a_i\alpha_i^* = a_0 \mu_{\cdot 1} + \sum_1^p a_i(\mu_{i1}-\mu_{\cdot 1})$ is estimable, then $a_0 x_{\cdot 1} + \sum_1^p a_i(x_{i1}-x_{\cdot 1})$ is its unbiased estimate and is obtained by replacing μ^*, α_i^*'s by their unique least squares estimates.

Thus in the case of the linear model (3.95) where the design matrix is not of full rank and hence the parameters do not have unique estimates, we can, with an added restriction $\sum_1^p \alpha_i = 0$, obtain their unique estimates.

The model

$$(X_{ij}) = \mu^* + \alpha_i^* , \quad \sum_1^p \alpha_i^* = 0 \tag{3.96}$$

is called a reparametrization of the original model given in (3.94).

To analyse a linear model which is not of full rank, we will reduce it to a full rank model by means of reparametrization and then use the results obtained in Sections 3.2-3.4.

By a reparametrization of the linear model $E(\underline{X}) = X\underline{\beta}$ we obviously mean a transformation of $\underline{\beta}$ to $\underline{\alpha} = A\underline{\beta}$ where A is a matrix of real elements such that each element of $\underline{\alpha}$ is an estimable function. We now establish the existence of such a reparametrization.

3.6.1. Existence of Reparametrized Models: Let us now consider a linear model $E(\underline{X}) = X\underline{\beta}$, $\mathrm{Cov}(\underline{X}) = \sigma^2 I$, where X is an N × p matrix of rank $k < p$, $\underline{\beta} = (\beta_1,\ldots,\beta_p)'$ and the N-vector $\underline{X}$ has a multivariate normal distribution. Since X'X is a positive semidefinite matrix of rank k, there exists a nonsingular matrix C of dimension p × p such that

$$C'X'XC = \begin{pmatrix} D & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}, \tag{3.97}$$

where D is a k × k matrix of rank k. Writing $C = (C_1, C_2)$ where C_1 is a p × k matrix containing the first k columns of C, from (3.95) we conclude that $C_1'X'XC_1 = D$, $C_2'X'XC_2 = \underline{0}$. Since D is of rank k, the rank of $C_1'X_1$ is k and $C_2'X' = \underline{0}$. Thus,

$$E(\underline{X}) = XCC^{-1}\underline{\beta} = (XC_1)\, U_1\underline{\beta}, \tag{3.98}$$

where $C^{-1} = (U_1, U_2)'$, and U_1 is a k × p matrix containing the first k rows of C^{-1}. Now writing

$\underline{\alpha} = U_1\underline{\beta}$, $XC_1 = Z$, we get

$$E(\underline{X}) = Z\underline{\alpha}, \quad \mathrm{Cov}(\underline{X}) = \sigma^2 I, \tag{3.99}$$

where Z is an N × k matrix of full rank and $\underline{\alpha}$ is a k-vector each of whose components is estimable. Note that the unique least squares estimate (which is also the maximum likelihood estimate) of α is given by

$$\bar{\underline{\alpha}} = (Z'Z)^{-1}Z'\underline{x}.$$

There may be other full rank reparametrizations of the same non-full rank model. The following argument shows that any two full rank reparametrizations are linearly related and hence the estimates of any estimable linear parametric functions from any two full rank reparametrization models are identical. Suppose $E(\underline{X}) = X\underline{\beta}$, where the rank of $X = k < p$, has two full rank reparametrizations $E(\underline{X}) = Z\underline{\alpha}$ and $E(\underline{X}) = W\underline{\gamma}$, where Z, W are two different N × k matrices of rank k. Since they are reparametrization models there exist matrices A, B of dimension k × p each such that $\underline{\alpha} = A\underline{\beta}$, $\underline{\gamma} = B\underline{\beta}$. Since the vectors $\underline{\alpha}, \underline{\gamma}$ are of the same dimension, there exists a nonsingular matrix E such that A = EB. In other words, any two full rank reparametrizations are linearly related.

Example 3.6. Consider the one-way classification with p = 2 and $N_1 = N_2 = 2$, with the linear model

$$E(X_{ij}) = \mu + \alpha_i \,, \quad j = 1,2;\ i = 1,2.$$

Let $\beta_i = \mu + \alpha_i$, $\underline{\beta} = (\mu, \alpha_1, \alpha_2)'$, $\underline{\alpha} = (\beta_1, \beta_2)'$. Then $\underline{\alpha} = U_1\underline{\beta}$, where

$$U_1 = \begin{pmatrix} 1,1,0 \\ 1,0,1 \end{pmatrix}.$$

Since $C^{-1} = (U_1, U_2)'$ is nonsingular, we get

$$C^{-1} = \begin{pmatrix} 1,1,0 \\ 1,0,1 \\ 0,1,1 \end{pmatrix}, \quad \text{so,} \quad C = \begin{pmatrix} \tfrac{1}{2}, & \tfrac{1}{2}, & -\tfrac{1}{2} \\ \tfrac{1}{2}, & -\tfrac{1}{2}, & \tfrac{1}{2} \\ -\tfrac{1}{2}, & \tfrac{1}{2}, & \tfrac{1}{2} \end{pmatrix}.$$

As

$$X = \begin{pmatrix} 1, & 1, & 0 \\ 1, & 0, & 0 \\ 1, & 0, & 1 \\ 1, & 0, & 1 \end{pmatrix}, \qquad C_1 = \begin{pmatrix} \tfrac{1}{2}, & \tfrac{1}{2} \\ \tfrac{1}{2}, & -\tfrac{1}{2} \\ -\tfrac{1}{2}, & \tfrac{1}{2} \end{pmatrix}$$

we get,

$$Z = XC_1 = \begin{pmatrix} 1,0 \\ 1,0 \\ 0,1 \\ 0,1 \end{pmatrix}.$$

The normal equations for the reparametrized model is given by $(Z'Z)\underline{\alpha} = Z'\underline{x}$. From this $\bar{\alpha}_1 = x_{1.}$, $\bar{\alpha}_2 = x_{2.}$. Now $\alpha_1 - \alpha_2$ is estimable and is equal to $\beta_1 - \beta_2$. Thus $\bar{\beta}_1 - \bar{\beta}_2 = x_{1.} - x_{2.}$.

On reparametrizing a linear model to the full rank model $E(\underline{X}) = Z\underline{\alpha}$, if the full rank design matrix Z is such that Z'Z is a diagonal matrix then it is called an orthogonal reparametrization.

One advantage in such orthogonal reparametrization is that the components of $\underline{\bar{\alpha}}$ are uncorrelated. The fact that such an orthogonal reparametrization always exists can be established in the same way as in the case of full rank reparametrization. Here we need to take C to be a k × k orthogonal matrix such that the matrix D is a diagonal matrix whose diagonal elements are the characteristic roots of X'X. Since C is orthogonal we get $\alpha = C_1'\underline{\beta}$. Furthermore, since

$$\underline{\bar{\alpha}} = (Z'Z)^{-1}Z'\underline{X} ,$$

we get

$$\text{Cov}(\underline{\bar{\alpha}}) = \sigma^2(Z'Z)^{-1} = \sigma^2 D^{-1} .$$

Thus $\bar{\alpha}_i$, $\bar{\alpha}_j$, $(i \neq j)$ are uncorrelated.

3.6.2. Estimation and Testing in the Reparametrized Model: From Theorem 3.14 it follows that, for the reparametrization model $E(\underline{X}) = Z\underline{\alpha}$, $\text{Cov}(\underline{X}) = \sigma^2 I$, where $\underline{X}$ has an N-variate normal distribution, the unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{N-k}(\underline{X} - Z\hat{\underline{\alpha}})'(\underline{X} - Z\hat{\underline{\alpha}}) . \qquad (3.100)$$

and $\hat{\underline{\alpha}}$ $(= \bar{\underline{\alpha}})$ has a k-variate normal distribution with mean $\underline{\alpha}$ and covariance matrix $\sigma^2(Z'Z)^{-1}$, $(N-1)\hat{\sigma}^2/\sigma^2$ is distributed as chisquare with N-k degrees of freedom, independently of $\hat{\underline{\alpha}}$. We can also write $\hat{\sigma}^2$ in terms of $\hat{\underline{\beta}}$ as

$$\hat{\sigma}^2 = \frac{1}{N-k}(\underline{X} - X\hat{\underline{\beta}})'(\underline{X} - X\hat{\underline{\beta}}) \qquad (3.101)$$

where $\hat{\underline{\beta}}$ is a solution of the normal equation

$$(X'X)\hat{\underline{\beta}} = X'\underline{X} . \qquad (3.102)$$

As we have seen in Section 3.2 that for any estimable parametric function $\underline{L}'\underline{\beta}$, the unique best linear unbiased estimator is given by $\underline{L}'\hat{\underline{\beta}}$, where $\hat{\underline{\beta}}$ is any solution of (3.102). Similarly we can conclude that $\hat{\sigma}^2$ is unique for any such solution $\hat{\underline{\beta}}$.

For testing the null hypothesis $H_0 : \underline{L}'\underline{\beta} = \delta$ where $\underline{L}'\underline{\beta}$ is an estimable parametric function, we reparametrize the linear model $E(\underline{X}) = X\underline{\beta}$ to the full rank mode $E(\underline{X}) = Z\underline{\alpha}$ such that $\alpha_1 = \underline{L}'\underline{\beta}$. Since under H_0, $t = (\hat{\alpha}_1-\delta)/\sqrt{c_{11}\hat{\sigma}^2}$, where $(Z'Z)^{-1} = (c_{ij})$ has Student t-distribution with N-k degrees of freedom, an optimum test for testing H_0 against alternatives $\underline{L}'\underline{\beta} \neq \delta$ is given by

$$\text{Reject } H_0 \text{ whenever } |t| \geq t_{1-\alpha/2,N-k} . \qquad (3.103)$$

Let us now consider the problem of testing the null hypothesis H_0: $\beta_1 = \dots = \beta_q = 0$, $q \leq k$ in the linear model $E(\underline{X}) = X\underline{\beta}$ where the rank of X is k.

Definition 3.2. Estimable Hypothesis: A hypothesis H_0 is called estimable if there exists linearly estimable functions $\underline{L}'_1\underline{\beta}, \dots, \underline{L}'_s\underline{\beta}$ such that H_0 is true if and only if $\underline{L}'_i\underline{\beta} = 0$, $i = 1, \dots, s$.

In testing hypotheses in linear models, where the design matrix is not of full rank, we need to restrict ourselves to hypothesis which are estimable in the above sense. Let us assume H_0 is estimable and H_0 is true if and only if the linearly independent parametric functions $\alpha_i = \underline{L}'_i\underline{\beta} = 0$, $i = 1, \dots, s$. We now reparametrize it to the full rank model $E(\underline{X}) = Z\underline{\alpha} = Z_1\underline{\alpha}_{(1)} + Z_2\underline{\alpha}_{(2)}$ where $\underline{\alpha}_{(1)} = (\alpha_1, \dots, \alpha_s)'$, $\underline{\alpha}_{(2)} = (\alpha_{s+1}, \dots, \alpha_k)'$ and $Z = (Z_1, Z_2)$, Z_1 being the $N \times s$ submatrix of Z containing its first s columns. Thus the problem reduces to case (b) of Section 3.4 with H_0: $\underline{\beta}_{(1)} = \underline{0}$. Hence the test statistics equivalent to rejecting H_0 if

$$\frac{N-k}{s}\,\frac{q_1}{q_2} \geq F_{1-\alpha,s,N-k},$$

where $q_2 = (\underline{x} - Z\hat{\underline{\alpha}})'(\underline{x} - Z\hat{\underline{\alpha}}) = (\underline{x} - X\hat{\underline{\beta}})'(\underline{x} - X\hat{\underline{\beta}})$

$q_1 + q_2 = (\underline{x} - Z_2\hat{\underline{\alpha}}_{(2)})'(\underline{x} - Z_2\hat{\underline{\alpha}}_{(2)})$,

with $Z'_2Z_2\hat{\underline{\alpha}}_{(2)} = Z'_2\underline{x}$.

It is easy to see that like $q_2, q_1 + q_2$ can also be obtained from the original model $E(\underline{X}) = X\underline{\beta}$. Write $\beta_i = \dots = \beta_q$ in $E(\underline{X}) = X\underline{\beta}$ and call it $E(\underline{X}) = Z_2\underline{\alpha}_{(2)}$. Obtain a solution of the normal equations $Z'_2Z_2\hat{\alpha}_{(2)} = Z'_2\underline{x}$ and compute $(\underline{x} - Z_2\hat{\underline{\alpha}}_{(2)})'(\underline{x} - Z_2\hat{\underline{\alpha}}_{(2)})$.

3.7. TWO-WAY CLASSIFICATION

3.7.1. Two-way Classification Without Interaction: Consider an experiment where I different varieties of jute are grown on I different plots of the same size in each of J different locations. All plots are equally treated with the same chemical fertilizer and water treatment. Such an experiment is called a two-factor experiment, the factors being location (A) and variety of jute (B). Different locations constitute different levels of A and different varieties constitute different levels of B. The observations (yield per plot expressed in some convenient unit) can be arranged in a two-way classified (I × J) table where the rows correspond to the levels of A and the columns correspond to the levels of B. The (i,j)th element x_{ij} represents the yield of ith variety of jute at the jth location. The corresponding random variables x_{ij} are assumed to be independently normally distributed with mean μ_{ij} and variance σ^2. Throughout this section the replacement of a subscript in a variable by a dot will indicate that the variable has been averaged with respect to that subscript. For example

$$x_{i.} = \frac{1}{J}\sum_{j=1}^{J} x_{ij}, \quad \mu_{.j} = \frac{1}{I}\sum_{i=1}^{I} \mu_{ij}, \quad \text{etc.}$$

Since there is only one observation per cell, the combined effect of the two factors A, B, which is generally called the "interaction" of A with B, cannot be measured and we shall assume it to be zero. Mathematically, the function f(a,b) of two variables a, b, where a represents the level of the factor A and b represents the level of the factor B, is called a function of no interaction if and only if there exists functions g(a) and h(b) such that f(a,b) = g(a) + h(b). Otherwise we say that the factors interact. Obviously for a function f(a,b) of no interaction $f(a_1,b) - f(a_2,b)$ is independent of b. Such no interaction functions are also called additive functions.

A suitable model for two-way classification without interaction is given by

$$E(X_{ij}) = \alpha'_i + \beta'_j, \quad i = 1,\ldots,I;\ j = 1,\ldots, J, \tag{3.104}$$

where α'_i denotes the mean effect of the ith level of the factor A and β'_j denotes the mean effect of the jth level of the factor B. Clearly (3.104) represents a linear model in the parameters $\underline{\beta} = (\alpha'_1,\ldots,\alpha'_I,\ \beta'_1,\ldots,\beta'_J)'$. The design matrix X is given by the coefficients of these parameters.

Table 3.8. The design matrix X

$\underline{Y}$	α_1	α_2	. .	α_I	β_1	β_2	. .	β_J
Y_{11}	1	0	. .	0	1	0	. .	0
Y_{12}	1	0	. .	0	0	1	. .	0
.	.	..	.	...	..	.	.	.
Y_{1J}	1	0	. .	0	0	0	. .	1
.	.	.	.	. .	.	.	.	.
Y_{J1}	0	0	. .	1	1	0	. .	0
Y_{I2}	0	0	. .	1	0	1	. .	0
. .	. .	. .	. .	. .	. .	. .	. .	. .
Y_{IJ}	0	0	. .	1	0	0	. .	1

It is a matrix of dimension $N \times (I + J)$, $(N = IJ)$, and is of rank $I + J - 1$ $(N > I + J)$. Obviously all linear parametric functions are not estimable. However, we can rewrite (3.104) as

$$E(X_{ij}) = \mu + \alpha_i + \beta_j,\quad i = 1,\ldots,I,\ j = 1,\ldots,J; \tag{3.105}$$

$$\sum_1^I \alpha_i = 0,\qquad \sum_1^J \beta_j = 0\ ,$$

where α's, β's, and μ are uniquely determined by μ_{ij} (not so for (3.104)) as $\alpha_i = \mu_{i.} - \mu_{..}$, $\beta_j = \mu_{.j} - \mu_{..}$, $\mu = \mu_{..}$. In terms of α'_i, β'_j, $\alpha_i = \alpha'_i - \alpha'_.$, $\beta_j = \beta'_j - \beta'_.$, $\mu = \alpha'_. + \beta'_.$. Our main interest in this model is to estimate contrasts $\sum_1^I \ell_i \alpha_i$, $\sum_1^J c_j \beta_j$ and test the hypotheses

$H_{\underline{\alpha}} : \alpha_1 = \dots = \alpha_I = 0$; $H_{\underline{\beta}} : \beta_1 = \dots = \beta_J = 0$.

The quantity α_i is called the "effect" of the ith level of the factor A and is defined to be the excess of the mean of the ith level of the factor A over the "general mean" $\mu_{..} = \mu$. Similarly β_j is called the effect of the jth level of the factor B. Writing $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where $\gamma_{ij} = (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})$, the quantity γ_{ij} is called the interaction of the ith level of A with the jth level of B. The linear model $E(X_{ij}) = \mu_{ij}$ is a full rank model in the parameters μ_{ij}'s and hence all linear parametric functions of μ_{ij}'s are estimable. Thus in particular all α_i, all β_j and all γ_{ij} are estimable (the least squares estimate of μ_{ij} is x_{ij}). But if we have one observation per cell and $\gamma_{ij} \neq 0$, there is no estimate of error variance σ^2.

Since $\sum_1^I \alpha_i = 0$, $\sum_1^J \beta_j = 0$, in the linear model (3.105) all linear parametric functions of the parameters μ, α_i's, β_j's are estimable. Furthermore since (3.105) is a reparametrized model of (3.104), the model (3.105) will lead to the same estimate of any estimable linear function as in the case of the original model. As in the case of one-way classification the only linear combinations of α_i's or of β_j's that are estimable are contrasts.

The least squares estimates $\bar{\mu}$, $\bar{\alpha}_i$, $\bar{\beta}_j$ (which are also the maximum likelihood estimates) of the parameters μ, α_i, β_j are obtained by minimizing

$$s^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \mu - \alpha_i - \beta_j)^2$$

with respect to these parameters. This gives

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \bar{\mu} - \bar{\alpha}_i - \bar{\beta}_j) = 0,$$

$$\sum_{i=1}^{I} (x_{ij} - \bar{\mu} - \bar{\alpha}_i - \bar{\beta}_j) = 0$$

$$\sum_{j=1}^{J} (x_{ij} - \bar{\mu} - \bar{\alpha}_i - \bar{\beta}_j) = 0 .$$

Since $\sum_1^I \alpha_i = 0$, $\sum_1^J \beta_j = 0$ we can use $\sum_1^I \bar{\alpha}_i = 0$, $\sum_1^J \bar{\beta}_j = 0$ for solving these normal equations and this gives

$$\bar{\mu} = x_{..}\,, \quad \bar{\alpha}_i = x_{i.} - x_{..}\,, \quad \bar{\beta}_j = x_{.j} - x_{..}\,.$$

Clearly these estimates are unique. The error sum of squares, which is the minimum value of s^2 for all possible values of the parameters μ, α_i's, β_j's, is given by

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 .$$

The degrees of freedom is clearly $IJ - (I + J + 1) = (I - 1)(J - 1)$. The sum of squares due to the hypothesis H_α is the minimum value of

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \mu - \beta_j)^2$$

with respect to the parameters μ, β_j's and is equal to

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - x_{.j})^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 + \sum_1^I J(x_{i.} - x_{..})^2 .$$

Hence the sum of squares due to the deviation from the hypothesis H_α or equivalently the sum of squares due to rows is given by $\sum_{i=1}^{I} J(x_{i.}-x_{..})^2$, with degrees of freedom $(IJ - J) - (I - 1)(J - 1) = I - 1$. Similarly with the sum of squares due to the deviation from the hypothesis $H_{\underline{\beta}}$ is $\sum_{j=1}^{J} I(x_{.j} - x_{..})^2$ with $J - 1$ degrees of freedom. Thus the likelihood ratio test of level α for testing $H_{\underline{\alpha}}$ is equivalent to rejecting it if

$$\frac{(I-1)(J-1)}{I-1} \; \frac{\sum_{i=1}^{I} J(x_{i.} - x_{..})^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2} \ge F_{1-\alpha, I-1, (I-1)(J-1)} \qquad (3.106)$$

and the likelihood ratio test of $H_{\underline{\beta}}$ is equivalent to ,

$$\frac{(I-1)(J-1)}{J-1} \frac{\sum_{j=1}^{J} I(x_{.j}-x_{..})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2} \geq F_{1-\alpha,J-1,(I-1)(J-1)} . \quad (3.107)$$

The reader will find it instructive to verify directly that these are likelihood ratio tests and the test statistics have the F-distribution under the respective null hypotheses.

The ANOVA for the two-way classification without interaction is given in Table 3.9. Note

$$\sigma_A^2 = \frac{1}{I-1}\sum_{1}^{I}\alpha_i^2 , \quad \sigma_B^2 = \frac{1}{J-1}\sum_{j=1}^{J}\beta_j^2 .$$

Table 3.9. ANOVA table for two-way classification without interaction

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	Expected mean sum of squares
Rows	$I-1$	$s_1^2 = \sum_{i=1}^{I} J(x_{i.}-x_{..})^2$	$\frac{s_1^2}{I-1}$	$\sigma^2+J\sigma_A^2$
Columns	$J-1$	$s_2^2 = \sum_{j=1}^{J} I(x_{.j}-x_{..})^2$	$\frac{s_2^2}{J-1}$	$\sigma^2+I\sigma_B^2$
Error	$(I-1)(J-1)$	$s_3^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2$	$\frac{s_3^2}{(I-1)(J-1)}$	σ^2
Total	$IJ-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-x_{..})^2$		

If $H_{\underline{\alpha}}$ is rejected then it is necessary to divide the α_i's into groups such that α_i's belonging to the same group are equal and those belonging to different groups are different. This is done by means of the t-test or by multiple range tests as in the case of one-way classification. The rejection of $H_{\underline{\beta}}$ could be treated similarly.

In the terminology of Section 3.1, with the model (3.105), the hypothesis $H_{\underline{\alpha}}$ is a linear hypothesis with $r = I-1$ and $s = 1+(I-1) + (J-1) = I+J-1$ and $N-s = (I-1)(J-1)$. The least squares estimates of α_i, β_j, μ under π_Ω are given by (3.106) and

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-\hat{\mu}_{ij})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2 .$$

Under the hypothesis $H_{\underline{\alpha}} : \alpha_1 = \ldots = \alpha_I = 0$, the least squares estimates of β_j and μ are

$$\hat{\hat{\beta}}_j = x_{.j}-x_{..} \,, \quad \hat{\hat{\mu}} = x_{..}$$

and hence $\hat{\mu}_{ij} - \hat{\hat{\mu}}_{ij} = x_{i.}-x_{..}$.

Hence from (3.7) the test statistic w for testing $H_{\underline{\alpha}}$ is given by

$$w = \frac{(J-1)\sum_{i=1}^{I} J(x_{i.}-x_{..})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2} \tag{3.108}$$

and the distribution of W depends only on the parameter

$$\delta^2 = J\sum_{i=1}^{I}\frac{(\mu_{i.}-\beta_{..})^2}{\sigma^2} = J\sum_{i=1}^{I}\frac{\alpha_i^2}{\sigma^2} .$$

Thus the uniformly most powerful invariant test of $H_{\underline{\alpha}}$ against the alternatives $\delta^2 > 0$ is given by (3.106). Similarly the test given in (3.107) is uniformly most powerful invariant for testing $H_{\underline{\alpha}}$ against the alternatives

$$I\sum_{j=1}^{J}\beta_j^2/\sigma^2 > 0 .$$

Example 3.7. The following data constitute a part of those collected from an experiment with different crosses of Barley which were raised in Hissar during 1971-1972 in 4 locations. We are interested in testing the null hypothesis H_C that the cross means are all equal, and the null hypothesis H_L, that the location means are all equal. We assume that the data constitute a random sample from normal populations with the same variance σ^2.

Table 3.10. Average yield per plant in grams

	Locations			
Crosses	1	2	3	4
1	101.32	103.04	91.78	90.22
2	77.08	76.36	97.40	95.32
3	91.96	68.16	85.62	96.16
4	92.46	99.66	77.64	88.42
5	91.62	89.18	105.32	80.26
6	92.62	78.74	78.74	84.78
7	72.18	74.02	72.27	82.74
8	77.68	82.96	93.66	90.48
9	108.72	110.14	109.28	104.36
10	85.46	69.64	86.28	68.96

Table 3.11. ANOVA for Table 3.10

Sources of variations	Degrees of freedom	Sum of squares	F
Between Crosses	9	3284.6025	4.9445
Between Locations	3	110.4387	0.4987
Error	27	1192.8469	
Total	39	4587.8881	

Since $F_{\alpha,9,27} = 3.15$ for $\alpha = .01$ we reject H_c at $\alpha = .01$.

3.7.2. Two-way Classification with Interactions: Here we will consider the two-way classification with an equal number K of observations per cell. Let x_{ijk} denote the kth observation in the (i,j)th cell, $i = 1,\ldots,I$, $j = 1,\ldots,J$, $k = 1,\ldots,K$ and let x_{ijk} be independently distributed normal random variable with the same variance σ^2 and

$$E(X_{ijk}) = \mu_{ij} \ . \tag{3.109}$$

It is a linear model in parameters μ_{ij}'s. Its design matrix, given by the coefficients of μ_{ij}'s in (3.109) is a matrix of full rank. Hence all linear parametric functions of μ_{ij}'s are estimable.

In terms of the general mean μ, effects α_i, β_j and interactions γ_{ij} we can write

$$E(X_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}\,,$$

$$i = 1,\ldots,I,\ j = 1,\ldots,J,\ k = 1,\ldots,K,$$

$$\sum_1^I \alpha_i = 0,\quad \sum_1^J \beta_j = 0,\quad \sum_{i=1}^I \gamma_{ij} = 0,\quad \sum_{j=1}^J \gamma_{ij} = 0\,.$$

The hypothesis that we usually wish to test are

(i) $H_{\underline{\alpha}} : \alpha_1 = \ldots = \alpha_I = 0$,

(ii) $H_{\underline{\beta}} : \beta_1 = \ldots = \beta_J = 0$,

(iii) $H_{\underline{\gamma}}$: all $\gamma_{ij} = 0$.

The least squares estimates (which are also the maximum likelihood estimates) $\bar{\mu}$, $\bar{\alpha}_i$'s, $\bar{\beta}_j$'s and $\bar{\gamma}_{ij}$'s of μ, α_i's, β_j's and γ_{ij}'s are obtained by minimizing

$$s^2 = \sum_i \sum_j \sum_k (x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2$$

with respect to these parameters. Differentiating s^2 with respect to these parameters and equating these partial derivatives to zero we obtain a set of normal equations which, with

$$\sum_1^I \bar{\alpha}_i = 0,\quad \sum_1^J \bar{\beta}_j = 0,\quad \sum_{i=1}^I \bar{\gamma}_{ij} = 0,\quad \sum_{j=1}^J \bar{\gamma}_{ij} = 0$$

give us the unique estimates,

$$\bar{\mu} = x_{\ldots},\quad \bar{\alpha}_i = x_{i..} - x_{\ldots},\quad \bar{\beta}_j = x_{.j.} - x_{\ldots},$$

$$\bar{\gamma}_{ij} = (x_{ij.} - x_{i..} - x_{.j.} + x_{\ldots})\,.$$

The error sum of squares which is the minimum value of s^2 with respect to these parameters is given by

$$\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2$$

with IJK-IJ=IJ(K-1) degrees of freedom. The sum of squares due to the hypothesis $H_{\underline{\alpha}}$ is given by the minimum value of

$$\sum_i \sum_j \sum_k (x_{ijk} - \mu - \beta_j - \gamma_{ij})^2$$

with respect to the parameters μ, β_j's and γ_{ij}'s, and is given by

$$\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2 + JK \sum_i (x_{i..} - x_{...})^2 ,$$

with IJ(K-1) + I-1 degrees of freedom. Hence the sum of squares due to the deviation from the hypotheses, which is also called the sum of squares due to the effect of A is given by

$$JK \sum_i (x_{i..} - x_{...})^2 ,$$

with I-1 degrees of freedom. Similarly the sum of squares due to the effect of B is given by

$$IK \sum_j (x_{.j.} - x_{...})^2$$

with J-1 degrees of freedom. The sum of squares due to the hypothesis $H_{\underline{\alpha}}$ is the minimum value of

$$\sum_i \sum_j \sum_k (x_{ijk} - \mu - \alpha_j - \beta_j)^2 ,$$

with respect to these parameters μ, α_i's, β_j's, and is equal to

$$\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2 + K \sum_i \sum_j (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2$$

with IJ(K-1) + (I-1)(J-1) degrees of freedom. Hence the sum of squares due to the deviation from the hypothesis $H_{\underline{\gamma}}$ is given by

$$K \sum_i \sum_j (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2 ,$$

with (I-1)(J-1) degrees of freedom. It may be remarked that except for the parameters which are zero under $H_{\underline{\alpha}}$, $H_{\underline{\beta}}$ and $H_{\underline{\gamma}}$, respectively the least squares estimates under these hypotheses are the same as under Ω (the parametric space for this problem).

The likelihood ratio test of level α for testing $H_{\underline{\alpha}}$ is thus equivalent to rejecting $H_{\underline{\alpha}}$ whenever

$$\frac{IJ(K-1)}{I-1}\;\frac{JK\sum_{1}^{I}(x_{i..}-x_{...})^2}{\sum_i\sum_j\sum_k(x_{ijk}-x_{ij.})^2} \ge F_{1-\alpha,I-1,IJ(K-1)} \qquad (3.110)$$

The likelihood ratio test of level α for testing $H_{\underline{\beta}}$ is equivalent to rejecting $H_{\underline{\beta}}$ whenever

$$\frac{IJ(K-1)}{J-1}\;\frac{IK\sum_{1}^{J}(x_{.j.}-x_{...})^2}{\sum_i\sum_j\sum_k(x_{ijk}-x_{ij.})^2} \ge F_{1-\alpha,J-1,IJ(K-1)}\;; \qquad (3.111)$$

and that of $H_{\underline{\gamma}}$ is equivalent to rejecting $H_{\underline{\gamma}}$ whenever

$$\frac{IJ(K-1)}{(I-1)(J-1)}\;\frac{K\sum_i\sum_j(x_{ij.}-x_{i..}-x_{.j.}+x_{...})^2}{\sum_i\sum_j\sum_k(x_{ijk}-x_{ij.})^2}$$

$$\ge F_{1-\alpha,(I-1)(J-1),IJ(K-1)}\;, \qquad (3.112)$$

The reader can verify these facts directly as in the case of one-way classification without any difficulties.

Let

$$\sigma_A^2 = \frac{1}{I-1}\sum_i \alpha_i^2,\quad \sigma_B^2 = \frac{1}{J-1}\sum_j \beta_j^2,\quad \sigma_{AB}^2 = \frac{1}{(I-1)(J-1)}\sum_i\sum_j \gamma_{ij}^2 \;.$$

If the hypothesis $H_{\underline{\alpha}}$ or $H_{\underline{\beta}}$ is rejected one can use the t-test or the multiple range test to find out which pair of the α_i's or which pair of the β_j's are significantly different. If $H_{\underline{\gamma}}$ is rejected then one would not ordinarily explore it further, although theoretically t and multiple range tests can be applied to find out the interactions which are significantly different from zero.

In the same way as in the case of two-way classification without interaction it is easy to verify that the test given in (3.110) is uniformly most powerful invariant for testing $H_{\underline{\alpha}}$ against the alternatives $JK\sum_{i=1}^{I}\alpha_i^2/\sigma^2 > 0$. Note that here $r=I-1$, $s = IJ$, $N-s=(K-1)IJ$.

Table 3.12. ANOVA table for the two-way classification with interaction

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	Expected sum of squares
Effects of A	I-1	$s_1^2 = JK \sum_i (x_{i..}-x_{...})^2$	$\frac{s_1^2}{I-1}$	$\sigma^2+JK\sigma_A^2$
Effects of B	J-1	$s_2^2 = IK \sum_j (x_{.j.}-x_{...})^2$	$\frac{s_2^2}{J-1}$	$\sigma^2+IK\sigma_B^2$
Interaction AB	(I-1)(J-1)	$s_3^2 = \sum_i \sum_j \sum_k (x_{ij.}-x_{i..}-x_{.j.}-x_{...})^2$	$\frac{s_3^2}{(I-1)(J-1)}$	$\sigma^2+K\sigma_{AB}^2$
Error	IJ(K-1)	$s_4^2 = \sum_i \sum_j \sum_k (x_{ijk}-x_{ij.})^2$	$\frac{s_4^2}{IJ(K-1)}$	σ^2
Total	IJK-1	$\sum_i \sum_j \sum_k (x_{ijk}-x_{...})^2$		

Similar results hold for the tests given in (3.111) and (3.112).

3.8. THREE-WAY CLASSIFICATION WITH INTERACTIONS

Suppose there are three factors A,B,C affecting the observations, the factor A has I levels, the factor B has J levels and the factor C has K levels. We shall assume that there is an equal number of M observations for each treatment combination (i,j,k) where i denotes the level of the factor A, j denotes the level of the factor B and k denotes the level of the factor C. Let x_{ijkm} denote the mth observation on the treatment combination (i,j,k); i = 1,...,I, j = 1,...,J, k = 1,...,K, m = 1,...,M and let x_{ijkm}'s be independently distributed normal random variables with mean μ_{ijk} and the same variance σ^2.

For each fixed level k of the vector C

$$\mu_{ijk}-\mu_{i.k}-\mu_{.jk}+\mu_{..j}$$

is the interaction of the ith level of A with the jth level of B.

If we take the average of this quantity for all levels of C we get

$$\alpha^{AB}_{ij} = (\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...})$$

and it is called the interaction effect of the ith level of A with the jth level of B. Similarly, the interaction α^{AC}_{ik} of the ith level of A with the kth level of C and the interaction α^{BC}_{jk} of the jth level of B with the kth level of C are given by

$$\alpha^{AC}_{ik} = (\mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu_{...}),$$

$$\alpha^{BC}_{jk} = (\mu_{.jk} - \mu_{..k} - \mu_{.j.} + \mu_{...}) .$$

The effect α^{A}_{i} of the ith level of A is obtained by averaging for all levels of B and C and is given by

$$\alpha^{A}_{i} = (\mu_{i..} - \mu_{...}) .$$

Similarly we define

$$\alpha^{B}_{j} = (\mu_{.j.} - \mu_{...}), \quad \alpha^{C}_{k} = (\mu_{..k} - \mu_{...}) .$$

The quantity $\mu_{...}$ is defined as the general mean. As there are three factors it is also necessary to have a measure of the combined effect of them. Let

$$\alpha^{ABC}_{ijk} = (\mu_{ijk} - \mu_{i.k} - \mu_{.jk} + \mu_{..k}) - (\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...}) .$$

Obviously it gives us a measure of how the interaction of the ith level of A with the jth level of B when the factor C is at the fixed level k differs from its average value when it is averaged for all levels of C. There is no dependence on C if and only if $\alpha^{ABC}_{ijk} = 0$. The quantity α^{ABC}_{ijk} is called the three factor interaction between the ith level of A, the jth level of B and the kth level of C.

With these quantities we could write

$$E(X_{ijkm}) = \mu + \alpha^{A}_{i} + \alpha^{B}_{j} + \alpha^{C}_{k} + \alpha^{AB}_{ij} + \alpha^{BC}_{jk} + \alpha^{AC}_{ik} + \alpha^{ABC}_{ijk}$$

$i = 1,\ldots,I; \quad j = 1,\ldots.J; \quad k = 1,\ldots,K$

$$\sum_i \alpha_i^A = 0, \quad \sum_j \alpha_j^B = 0, \quad \sum_k \alpha_k^C = 0 ,$$

$$\sum_i \alpha_{ij}^{AB} = 0, \quad \sum_j \alpha_{ij}^{AB} = 0, \quad \sum_j \alpha_{jk}^{BC} = 0, \quad \sum_k \alpha_{jk}^{BC} = 0 ,$$

$$\sum_i \alpha_{ik}^{AC} = 0, \quad \sum_k \alpha_{ik}^{AC} = 0, \quad \sum_i \alpha_{ijk}^{ABC} = 0 ,$$

$$\sum_j \alpha_{ijk}^{ABC} = 0, \quad \sum_k \alpha_{ijk}^{ABC} = 0 .$$

As in the two-way classification, the hypotheses of main interest are all effects of each factor are zero, all two-factor interactions are zero and the three-factor interaction is zero. For example, for effect of A

$$H_A : \alpha_1^A = \dots = \alpha_I^A = 0 .$$

Similarly, for the interaction of A, B and for the interaction of A, B, C

$$H_{AB}: \text{all } \alpha_{ij}^{AB} = 0$$

$$H_{ABC}: \text{all } \alpha_{ijk}^{ABC} = 0 .$$

The sum of squares due to the deviation from each of these hypotheses and the error sum of squares can be obtained in the same way as in the case of two-way classification.

The least squares estimates (which are the same as the maximum likelihood estimates) of these parameters are given by

$$\bar{\mu} = x_{...}, \quad \bar{\alpha}_i^A = x_{i..} - x_{...}, \quad \bar{\alpha}_{ij}^{AB} = (x_{ij.} - x_{i..} - x_{.j.} + x_{...}),$$

$$\bar{\alpha}_{ijk}^{ABC} = (x_{ijk.} - x_{ij..} - x_{i.j.} - x_{.jk.} + x_{i...} + x_{.j..} + x_{..k.} - x_{....})$$

and similar expressions for other effects and interactions. The sum of squares due to the deviations form various hypotheses and the error sum of squares are given in Table 3.13.

Table 3.13. ANOVA table for the three-way classification with interaction

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	Expected mean sum of squares
Effects of A	$I-1$	$s_1^2 = JKM \sum_i (\bar{\alpha}_i^{A})^2$	$\frac{s_1^2}{I-1}$	$\sigma^2+JKM\sigma_A^2$
Effects of B	$J-1$	$s_2^2 = IKM \sum_j (\bar{\alpha}_j^{B})^2$	$\frac{s_2^2}{J-1}$	$\sigma^2+IKM\sigma_B^2$
Effects of C	$K-1$	$s_3^2 = IJM \sum_k (\bar{\alpha}_k^{C})^2$	$\frac{s_3^2}{K-1}$	$\sigma^2+IJM\sigma_C^2$
Interaction AB	$(I-1)(J-1)$	$s_4^2 = KM \sum_i \sum_j (\bar{\alpha}_{ij}^{AB})^2$	$\frac{s_4^2}{(I-1)(J-1)}$	$\sigma^2+ KM\sigma_{AB}^2$
Interaction AC	$(I-1)(K-1)$	$s_5^2 = JM \sum_i \sum_k (\bar{\alpha}_{ik}^{AC})^2$	$\frac{s_5^2}{(I-1)(K-1)}$	$\sigma^2+JM\sigma_{AC}^2$
Interaction BC	$(J-1)(K-1)$	$s_6^2 = IM \sum_j \sum_k (\bar{\alpha}_{jk}^{BC})^2$	$\frac{s_6^2}{(J-1)((K-1)}$	$\sigma^2+IM\sigma_{BC}^2$
Interaction ABC	$(I-1)(J-1)(K-1)$	$s_7^2 = M \sum_i \sum_j \sum_k (\bar{\alpha}_{ijk}^{ABC})^2$	$\frac{s_7^2}{(I-1)(J-1)(K-1)}$	$\sigma^2+M\sigma_{ABC}^2$
Error	$IJK(M-1)$	$s^2=\sum_i \sum_j \sum_k \sum_m (y_{ijkm}-y_{ijk.})^2$	$\frac{s^2}{IJK(M-1)}$	σ^2
Total	$IJKM-1$	$\sum_i \sum_j \sum_k \sum_m (y_{ijkm}-y_{....})^2$		

where

$$\sigma_A^2 = \frac{1}{I-1} \sum_i (\alpha_i^A)^2 ,$$

$$\sigma_{AB}^2 = \frac{1}{(I-1)(J-1)} \sum_i \sum_j (\alpha_{ij}^{AB})^2 ,$$

$$\sigma_{ABC}^2 = \frac{1}{(I-1)(J-1)(K-1)} \sum_i \sum_j \sum_k (\alpha_{ijk}^{ABC})^2$$

and similar notations for σ_B^2 , σ_C^2, σ_{BC}^2 and σ_{AC}^2 . The likelihood ratio test of level α of any hypothesis, say H_{AB} is equivalent to rejecting H_{AB} if

$$\frac{IJK(M-1)}{(I-1)(J-1)} \frac{s_4^2}{s^2} \geq F_{1-\alpha,(I-1)(J-1);IJK(M-1)} ,$$

and similarly for other hypotheses. We leave it as an exercise to the reader to verify the uniformly most powerful invariant property of these tests.

The case of higher-way classification is too complicated to reproduce here.

EXERCISE 3

1. For the linear model (1.3.9) with rank $X = r \leq p$, show that at most r linearly independent linear parametric functions are estimable.

2. Let $X_1, \ldots, X_6$ be independent random variables having a common variance σ^2 and expectations given by

$$E(X_1) = \beta_1 + \beta_5, \ E(X_2) = \beta_2 + \beta_5$$

$$E(X_3) = \beta_3 + \beta_6, \ E(X_4) = \beta_4 + \beta_6$$

$$E(X_5) = \beta_1 + \beta_3, \ E(X_6) = \beta_2 + \beta_4$$

Show that $\beta_1 - \beta_2$ is estimable and find at least three unbiased estimates of $(\beta_1-\beta_2)$. Find also its best linear unbiased estimator and find its variance. Establish further that $\beta_1+\beta_2$ is not estimable. Obtain an unbiased estimate of σ^2.

3. Consider the linear model given in (1.3.9). Find the p-vector $\underline{\ell}$ such that $\underline{\ell}'\beta$ is estimable and the ratio of variance of its least square estimator to $\underline{\ell}'\underline{\ell}$ is maximum.

4. For the linear model (3.9) show that any linear function $\underline{L}'\underline{X}$ is the best linear unbiased estimator of $E(\underline{L}'\underline{X})$ if and only if $Cov(\underline{L}'\underline{X}, \underline{b}'\underline{X}) = 0$ for all N-vectors $\underline{b}$ satisfying $E(\underline{b}'\underline{X}) = 0$.

5. For the linear model (3.9) define the p linear functions $\phi_1, \ldots, \phi_p$, where

$\phi_j = \alpha_j'\underline{X}$, $j = 1, \ldots, p$,

$\underline{\alpha}_j$ being the jth column vector of X.

(i) Show that if the linear function

$\ell_1\phi_1 + \ldots + \ell_p\phi_p$

has zero expectation then it is identically equal to zero.

(ii) Show that every non-trivial linear function $\ell_1\phi_1 + \ldots + \ell_p\phi_p$ is the best linear unbiased estimator of its expectation.

6. Let $X_1, \ldots, X_N$, $N > p + 1$, be independent random variables with

$E(X_i) = \alpha + \beta_1(x_{1i} - \bar{x}_1) + \ldots + \beta_p(x_{pi} - \bar{x}_p)$

$V(X_i) = \sigma^2$, $i = 1, \ldots, N$,

~~Here,~~ Where x_{ij} are known constants,

and $\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{ji}$.

Express this as a linear model as given in (3.9) and assuming that X is of full rank obtain the best linear unbiased estimators of α, $\beta_1, \ldots, \beta_p$ and their variances and covariances. Assuming that $\underline{X}$ has an N-variate normal distribution, find suitable tests of the following:

(i) $H_0 : \alpha = \alpha_0$,   (ii) $H_0 : \beta_1 = \beta_1^0, \ldots, \beta_p = \beta_p^0$,

(iii) $H_0 : \beta_i = \beta_j$ $(i \neq j)$,   (iv) $H_0 : \beta_1 = \ldots = \beta_p$

where $\beta_1^0, \ldots, \beta_p^0$ are specified constants.

7. Consider the linear model (3.9) with

~~$(\underline{X}) = X\underline{\beta}$~~ , $Cov(\underline{Y}) = \sigma^2 V$

$E(\underline{X}) = X\underline{\beta}$

where V is a known positive definite matrix of dimension N × N. Prove that the best linear unbiased estimator of an estimable parametric function $\underline{L}'\underline{\beta}$ is given by $\underline{L}'\underline{\bar{\beta}}$ where $\underline{\bar{\beta}}$ is any solution of

$$X'V^{-1}X\underline{\beta} = X'V^{-1}\underline{X} .$$

8. Prove that the result of problem 4 also holds for the linear model assumed in problem 7.

9. As an application of the result mentioned in problem 8 show that if

$$E(X_r) = \mu + \sigma \sum_{i=1}^{r} (N-i+1)^{-1} , \quad r = 1,\dots,N$$

$$\mathrm{Cov}(X_r,X_s) = \sigma^2 \sum_{i=1}^{r} (N-i+1)^{-2}, \quad 1 \le r \le s \le N ,$$

then the best linear unbiased estimators of μ and σ are given by,

$$\hat{\mu} = \frac{NX_1-\bar{X}}{N-1} \quad \text{and} \quad \hat{\sigma} = \frac{N\bar{X}-X_1}{N-1} ,$$

where $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$.

10. Let $\underline{X} = (X_1,\dots,X_N)'$ be such that

$E(\underline{X}) = \underline{\mu}$, $\mathrm{Cov}(\underline{X}) = \Sigma$.

Show that

$E(\underline{X}'A\underline{X}) = \underline{\mu}'A\underline{\mu} + \mathrm{tr}(A\Sigma)$.

11. For the linear model (3.9) assume that $\underline{X}$ has a multivariate normal distribution and X is of rank p. Prove that

$$1-\alpha=p\{|\underline{\ell}'(\underline{\bar{\beta}}-\underline{\beta})| \le (pF_{1-\alpha,p,N-p})^{\frac{1}{2}} (\bar{\sigma}) (\underline{\ell}'(X'X)^{-1}\underline{\ell})^{\frac{1}{2}} \text{ for all } \underline{\ell} \ne \underline{0}\} .$$

12. In (3.93) show that under $H_0:\theta_1 = \dots = \theta_p$, U is distributed as a chisquare with p-1 degrees of freedom.

13. For the linear model (fixed effect)

$$E(X_i) = \beta_1 x_{1i} + \beta_2 x_{2i} , \quad i = 1,\dots,N,$$

$$\mathrm{Cov}(X_i) = \sigma^2 ,$$

where $X_1,\dots,X_N$ are independently normally distributed the following data were obtained.

y_i	6.0	13.0	13.0	29.2	33.1	32.0	46.2
x_{1i}	1	2	3	4	5	6	8
x_{2i}	10	10	12	11	15	15	18

(i) Find the least squares estimates of these parameters;

(ii) Test the null hypothesis $H_0 : \beta_1 = 0$.

14. For the two-way classification without interaction find the following:

(i) the 95% confidence interval of $\alpha_i - \alpha_{i'}$, $(i \neq i')$

(ii) the 95% confidence interval of $\beta_j - \beta_{j'}$, $(j \neq j')$.

15. Table 3.14 gives the birth weight (in lbs) of six litters of pigs, the litters being of different sizes. Analyse the following data to examine if litter size is an assignable cause of variation in birth weight.

Table 3.14. Birth weight in lbs of pigs by litter sizes

litter size	1	2	3	4	5	6
weight	3.5	3.2	2.6	3.1	2.6	2.9
	3.2	3.3	2.6	2.9	2.2	3.2
	3.8	3.2	2.9	3.1	2.0	2.1
	2.3	2.9	2.0	2.5	2.5	3.3
	1.7	2.5	2.0	1.2	1.2	2.8
	1.9	2.6	2.1	1.2	1.2	
	2.9	2.8				
		2.7				

16. Suppose we want to determine the individual weight of p objects by weighing them in a balance. One method is to weigh each one N times, and estimate its weight by the average of these N observations. If σ^2 is the error variance of each observation, then the precision of each estimate is σ^2/N. A second method is to weigh the objects in the following way: Place some of these objects in one pan and the rest in the other pan and place weight to achieve equilibrium. Let $x_{ij} = 0$, 1 or -1 according as the jth object is not used, placed on the right pan or placed on the left

pan respectively and let x_i denotes the weight required to gain equilibrium in the ith weighing. Then

$$x_{i1}\omega_1 + \ldots + x_{ip}\omega_p = x_i ,$$

where ω_k is the true weight of the kth object. Suppose this weighing process is repeated N times $(i = 1,\ldots,N)$ using different combinations of the objects for different pans. Thus we get a system of N linear equations in $\underline{\omega} = (\omega_1,\ldots,\omega_p)'$, from which ω can be estimated by means of least squares. The design matrix X is the N × p matrix whose elements are either 0 or 1 or −1. Show that maximum precision of the estimates of these individual weights can be obtained if the elements of X are either 1 or −1 and the columns of X are orthogonal.

17. Four objects $0_1, 0_2, 0_3$ and 0_4 whose individual weights are to be determined, are weighed in a common balance according to the following design. The symbol 1 indicates that the object was placed on the right pan and −1 indicates that the object was placed on the left pan.

0_1	0_2	0_3	0_4	weight
1	1	1	1	20.2
1	−1	1	−1	8.1
1	1	−1	−1	9.7
1	−1	−1	1	1.9
1	−1	1	−1	8.2
1	1	−1	−1	10.4
1	−1	−1	1	2.0
1	1	1	1	19.9

Assuming that the errors are independently normally distributed with zero mean and common variance σ^2,

(i) Find the maximum likelihood estimates of the weights of these four objects;

(ii) Find the estimate of σ^2.

18. The data in Table 3.15 relating to lactation yield (10 lb unit) of daughters of 5 sires are taken from the records of Kankrej herd maintained at the Institute of Agriculture, Anand, Indi

Table 3.15. Lactation yield (10 lbs units) of daughters

Daughter	Sires				
	1	2	3	4	5
	258	180	112	313	340
	178	236	269	374	247
	243	120	247	320	193
	178	140	198	532	235
	111	227	357	203	252
	233	140	327	269	280
	60	277	200	153	194
	262	179	190	228	169
	191		261	184	242
	138		235	383	253
	197		165		
	234				
	123				

Analyze the data to see if lactation yields for different sires are different.

19. For the maximization of phosphate potential in soils of Madras state, India, a scheme was initiated by the Indian Council of Agricultural research in Madras during 1961-1962. The details of one of the experiments conducted under the scheme are given below.

(a) Treatments applied

1. Control
2. Superphosphate 30 lbs per acre
3. Superphosphate 50 lbs per acre
4. Rock-phosphate 30 lbs per acre
5. Rock-phosphate 60 lbs per acre
6. Superphosphate 30 lbs and 5 tons of farm yard manure per acre
7. Superphosphate 60 lbs and 5 tons of farm yard manure per acre

8. Rockphosphate 30 lbs and 5 tons of farm yard manure per acre
9. Rockphosphate 30 lbs and 5 tons of farm yard per acre
10. Farm yard manure 5 tons per acre

(b) Plot size 20' x 6'. (c) crop sown Ragi.

Table 3.16. Yield of Ragi in lbs per plot

Replications	Treatments 1	2	3	4	5	6	7	8	9	10
1	5.62	5.00	4.75	4.00	5.50	3.50	5.50	6.12	5.12	5.00
2	4.50	5.50	4.75	4.25	4.75	5.50	4.50	4.75	5.75	5.37
3	3.39	4.00	4.50	5.00	4.50	4.50	6.00	5.00	5.25	5.75
4	3.50	3.50	5.00	4.00	4.50	3.70	5.00	4.75	3.75	5.50

Analyse the data as a two-way classification and draw all possible conclusions.

20. Let $(\underline{x}-\underline{a})'M(\underline{x}-\underline{a}) \leq 1$, with M a symmetric p × p positive definite matrix, be a ellipsoid with centre $\underline{a}$. Show that for any $\underline{h} = (h_1,\dots,h_p)' \neq \underline{0}$, the point $\underline{y} = (y_1,\dots,y_p)'$ lies between two planes of support of the ellipsoid orthogonal to $\underline{h}$ if and only if $|h'(\underline{y}-\underline{a})| \leq (h'M^{-1}\underline{h})^{\frac{1}{2}}$.

21. Source: Statistical analysis in Psychology and Education by George A. Ferguson, McGraw Hill series in Education p.253.

The data given in Table 3.17 were obtained from an experiment designed to study the effect of two variables on measures of performance of rats in a maze test with three strains of rats viz bright, mixed and dull. A group of 8 rats from each strain was reared under free or restricted environmental conditions.

Table 3.17

Environment	Strain Bright	Mixed	Dull	Environment	Strain Bright	Mixed	Dull
Free	26, 14	41, 82	36, 87	Restricted	51, 35	39, 114	42, 133
	41, 16	26, 86	39, 99		96, 36	104, 92	92, 124
	28, 29	19, 45	59, 126		97, 28	137, 87	156, 68
	92, 31	59, 39	27, 104		22, 76	122, 64	144, 142

1. Analyze the data as a two-way layout under normal theory model.
2. Do the data show a significant interaction between two experimental variables?
3. Use T- and S-Methods to decide which pair of strains differ significantly.

22. Source: Engineering statistics by A.H. Bowker and S.Lieberman.

To study the effect of plate temperatures and filament lighting on transconductance of certain type of tube an experiment was performed with two different plate temperatures 550°F, 600°F and four different filament lighting current L_1, L_2, L_3 and L_4. Three replications of each combination of plate temperature and filament lighting current were used. The data recorded on transconductance measurements are given in Table 3.18.

Table 3.18

Plate temperature	Filament lighting current			
	L_1	L_2	L_3	L_4
550°F	3774	4710	4176	4540
	4369	4180	4140	4530
	4374	4514	4398	3964
600°F	4216	3828	4122	4484
	4524	4170	4280	4332
	4136	4180	4226	4390

Make a complete analysis of the data under normal theory assumption.

23. Source: Methods of statistical analysis, by C.H. Goulden, John Wiley, p.169.

The data in Table 3.19 give the volumes in milliters of loves of bread made under controlled condition from 100 gm batches of dough made with 12 different varieties of wheat flour and containing x = 0, 1, 2, 3, 4 milligrams of potassium bromate. Analyse the data as a two-way layout under normal theory model.

Table 3.19

Variety of wheat flour	Volume for x				
	0	1	2	3	4
1	950	1075	1055	975	880
2	890	980	955	865	825
3	830	850	820	770	735
4	770	815	765	725	700
5	860	1040	1065	975	945
6	835	960	985	915	845
7	795	900	905	880	785
8	800	860	870	850	850
9	750	940	1000	960	960
10	885	1000	1015	960	895
11	895	935	965	950	920
12	685	835	870	875	880
13	615	665	650	680	660
14	885	910	890	835	785
15	985	1075	1070	1015	1005
16	710	750	740	725	720
17	785	845	865	825	820

24. Source: Basic statistics for health science students by David S. Phillips, 1978, W.H. Freeman and Company, San Francisco.

Table 3.20 gives measurements on pulse rates taken from an experiment conducted to see if an exercise programme for CV patient had any effect on pulse rates. The pulse rates of 10 patients were measured before entering the programme and after being in the programme for 1,4 and 12 weeks. Analyze the data under normal theory model to see if the programme affected the pulse rate.

25. Treat the data presented in Table 3.19 as a one-way classification with 17 observations for each of the 4 population corresponding to 4 different values of x. Test under normal theory model the hypothesis of equality of variances by using the techniques of Section 3.5.3.

Table 3.20

Patient	Start	1 week	4 weeks	12 weeks
1	90	87	88	90
2	80	80	78	74
3	70	75	73	75
4	76	82	75	72
5	82	80	74	70
6	85	83	80	76
7	100	95	85	82
8	80	82	75	75
9	90	95	82	80
10	75	75	70	72

REFERENCES

Bose, R.C. (1944). The fundamental theorem of linear estimation, Proc. 31st Indian Science Congress. pp.203.

Chakravorty, M.C. (1962). Mathematics of design and analysis of experiments, Asia Publishing House, Bombay.

Duncan, D.B. (1955). Multiple range and multiple F tests, Biometrics, Vol. 2, pp.1-42.

Federer, W.T. (1955). Experimental designs, theory and applications, Macmillan Co., New York.

Giri, N. (1974). Introduction to probability and statistics, part I - probability, Marcel Dekker, N.Y.

Graybill, F.A. (1961). An introduction to linear statistical model, Vol.1, McGraw Hill, New York.

Hsu, P.L. (1938). Contribution to the theory of Student t-test as applied to the problems of two samples, Stat. Research Memoire, Vol.2, pp.1-24.

Kempthorne, O. (1952). Design and analysis of experiments, John Wiley and Sons, N.Y.

Keuls, M. (1952). The use of studentized range in connection with an analysis of variance. Fuplytica, Vol.1, pp.112-122.

Lehmann, E. (1959). Testing statistical hypotheses, John Wiley and Sons, N.Y.

Magness, T.A. and McGuire, M. (1962). Comparison of least squares and minimum variance estimates of regression parameters. Ann. Math. Statist., 33, pp.462-470.

Miller, R.G. Jr. (1966). Simultaneous statistical inference, McGraw Hill, N.Y.

Newman, D. (1939). The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation, Biometrika, 31, pp.20-30.

Rao, C.R. (1945a). Generalization of Markoff's theorem and tests of linear hypothesis, Sankhya, 7, pp. 9-15.

Rao, C.R. (1945b). Markoff's theorem with linear restrictions on parameters, Sankhya, 7, pp.16-19.

Rao, C.R. (1946). On the linear combination of observations and the general theory of least squares, Sankhya, 7, pp.237-256.

Rao, C.R. (1952). Advanced statistical methods in biometrical research, John Wiley and Sons, N.Y.

Rao, C.R. (1965). Linear statistical inference and its applications, John Wiley and Sons, N.Y.

Scheffe, H. (1953). A method of judging all contrasts in the analysis of variance, Biometrika, Vol.40, pp.87-104.

Scheffe, H. (1960). The analysis of variance, John Wiley and Sons, N.Y.

Student (1927), Errors of routine analysis, Biometrika 19, pp.151-164.

Tukey, J.W. (1951). Components in regression, Biometrics, Vol.7, pp.33-69.

Tukey, J.W. (1953). The problem of multiple comparisons, dittoed manuscripts of 396 pages. Princeton University.

CHAPTER 4. ANALYSIS OF COVARIANCE. OPTIMALITY OF DESIGNS

In general a scientific experiment is performed to know something that is unknown about a group of treatments and to test certain hypotheses about the corresponding treatment effects. The treatments are applied on experimental units selected for the experiment. Different scientific experiments have atleast one feature in common, that is, the variability of the experimental units. When variability of the experimental units is small relative to treatment differences and a complicated experimental design is not required, for the experiment, the experimenter simply takes a large number of observations, if feasible, on each treatment effect and computes their mean. The variation around the mean can be made as small as desired by taking more observations.

When there is considerable variation from observation to observation on the same treatment and it is not possible to take an unlimited number of observations, the techniques commonly used for reducing the variation are

1. Use of a proper experimental design,
2. Use of concomitant variables.

We shall discuss here these two techniques in the general framework. First we shall discuss the use of concomitant variables in linear models and their analysis. The analysis of linear models with concomitant variables is called the analysis of covariance. Then we shall discuss different criteria for selecting the optimum design in a given situation. Specific experimental designs along with proper allocation of experimental units and random allocation of treatments in blocks, their advantages and disadvantages and their analysis of variance and the analysis of covariance for some of them will be treated in Chapters 5 and 6.

If the above two techniques fail to control the experimental variability then the number of replications of different treatments (in other words the number of experimental units) will need to be increased to a point where adequate control of variability is attained.

4.1. ANALYSIS OF COVARIANCE

If, for given experimental material, the use of proper experimental design cannot control the experimental variation, the use of concomitant variables which are realted to the experimental variables may be effective in reducing the variability. Consider the one-way classification with

$$E(X_{ij}) = \beta_i,$$
$$\mathrm{Var}(X_{ij}) = \sigma^2, \; j=1,\ldots,N_i, \; i=1,\ldots,p.$$

If the ordinary analysis of variance for testing the hypothesis of equality of treatment effects shows a highly significant difference in treatment effects due to some factors affecting the experiment, taking the following model (which takes into account the effects of such factors)

$$E(X_{ij}) = \beta_i + \gamma t_{ij}$$
$$\mathrm{Var}(X_{ij}) = \sigma^2, \; j = 1,\ldots,N_i, \; i = 1,\ldots,p, \qquad (4.1)$$

where γ is the regression coefficient on the concomitant variables t_{ij}, (which are related to X_{ij}), the variability of treatment effects can be considerably reduced. In agricultural experiments with plots of land as experimental units the t_{ij} may be a measure of the fertility characteristic of the jth plot receiving the ith treatment. For experiments with animals to compare the growth rates of groups of animals receiving different diets, the observed differences in growth rates can be attributed to diet only if all animals are similar in some observable characters like weight, age, etc. which influence the growth rates. In the absence of similarity one can use t_{ij} as the weight or the age of jth animal receiving the ith treatment.

If in (4.1) one wished to consider the quadratic regression in t_{ij} instead of the linear one, that is,

$$E(X_{ij}) = \beta_i + \gamma_1 t_{ij} + \gamma_2 t_{ij}^2 ,$$
$$\mathrm{Var}(X_{ij}) = \sigma^2, \; j = 1,\ldots,N_i, \; i = 1,\ldots,p, \qquad (4.2)$$

the analysis of covariance with model (4.2) is the same as the analysis of covariance with two concomitant variables t_{ij}, t_{ij}^2.

In the case of the two-way classification of Section 3, with one observation per cell the linear model with concomitant variables is given by

$$E(X_{ij}) = \mu + \alpha_i + \beta_j + \gamma t_{ij}$$

or

$$E(X_{ij}) = \mu + \alpha_i + \beta_j + \gamma_1 t_{ij} + \gamma_2 \omega_{ij}, \tag{4.3}$$

where $i = 1,\ldots,I$, $j = 1,\ldots,J$ and $\sum_{i=1}^{I} \alpha_i = 0$, $\sum_{j=1}^{J} \beta_j = 0$, (x_{ij}, t_{ij}), $(x_{ij}, t_{ij}, \omega_{ij})$ are observations in the (i,j)th cell and t_{ij}, ω_{ij} are concomitant variables.

Though we shall consider the observations on concomitant variables as constants selected by the experimenter, in some experiments it may be more realistic to think these observations as values taken by some random variables. In such cases one has to treat $E(X_{ij})$ as the mean of the conditional distribution of the X_{ij} given the observed values of the concomitant variables. Obviously all distributions derived in Chapter 3 for statistical inferences with respect to this conditional linear model is also conditional as are the level of significance, the power and confidence coefficients. Since such an analysis requires discretion as to the joint distribution of the concomitant variables and has little applied interest, we shall not attempt to deal with it in this book. We shall now consider the analysis of covariance of one way classification, two-way classification with one observation per cell and the linear regression model. The analysis of covariance for some specific designs will be treated in Chapter 5. The analysis of covariance of higher-way classifications and the two-way classification with more than one observation per cell and other designs are left as exercises for the reader.

4.1.1. One Way Classification: We have already discussed the analysis of variance of a one-way classification with p normal populations in Chapter 3. We shall discuss here the analysis of covariance of such a classification with one concomitant variable. Let X_{ij}, $j = 1,\ldots,N_i$, $i = 1,\ldots,p$ be a random sample of size N_i from the ith normal population with

$$\mu_{ij} = E(X_{ij}) = \beta_i + \gamma t_{ij}$$
$$\mathrm{Var}(X_{ij}) = \sigma^2 \tag{4.4}$$

where the β_i, γ, σ^2 are unknown parameters and the t_{ij} are known constants (observations on a concomitant variable) and let

$$H_o\colon \beta_1 = \cdots = \beta_p = \beta(\text{say}). \tag{4.5}$$

Let

$$x_{i.} = \frac{1}{N_i}\sum_j x_{ij}, \quad x_{..} = \frac{1}{N}\sum_i\sum_j x_{ij},$$

$$x_{,j} = \frac{1}{p}\sum_i x_{ij}, \quad t_{i.} = \frac{1}{N_i}\sum_j t_{ij},$$

$$t_{.j} = \frac{1}{p}\sum_i t_{ij}, \quad t_{..} = \frac{1}{n}\sum_i\sum_i t_{ij},$$

$$N = \sum_i N_i. \tag{4.6}$$

We shall use these notations without any specific mention wherever necessary in this chapter. In the notation of Section 3.1, the parametric space π_Ω has dimension $s = p + 1$ and the parametric space π_ω under H_o has dimension $s - r = 2$. Under π_Ω one must minimize

$$\sum_{i,j}(x_{ij}-\mu_{ij})^2 = \sum_{i,j}(x_{ij}-\beta_i-\gamma t_{ij})^2, \tag{4.7}$$

which for fixed γ leads to $\beta_i = x_{i.} - \gamma t_{i.}$ and reduces the sum of squares to

$$\sum_i\sum_j (x_{ij} - x_{i.} - \gamma(t_{ij} - t_{i.}))^2. \tag{4.8}$$

Minimizing (4.8) with respect to γ we get

$$\hat{\gamma} = \frac{\sum_i \sum_j (x_{ij} - x_{i.})(t_{ij} - t_{i.})}{\sum_i \sum_j (t_{ij} - t_{i.})^2}$$

$$\hat{\beta}_i = x_{i.} - \hat{\gamma} t_{i.}, \quad \hat{\mu}_{ij} = \hat{\beta}_i + \hat{\gamma} t_{ij}, \tag{4.9}$$

Since

$$\begin{aligned} x_{ij} - \hat{\mu}_{ij} &= x_{ij} - \hat{\beta}_i - \hat{\gamma} t_{ij} \\ &= x_{ij} - x_{i.} - \hat{\gamma}(t_{ij} - t_{i.}), \end{aligned}$$

we get

$$\sum_i \sum_j (x_{ij} - \hat{\mu}_{ij})^2 = \sum_i \sum_j (x_{ij} - x_{i.})^2 - \frac{[\sum_i \sum_j (x_{ij} - x_{i.})(t_{ij} - t_{i.})]^2}{\sum_i \sum_j (t_{ij} - t_{i.})^2} \tag{4.10}$$

Under π_ω one minimizes

$$\sum_i \sum_j (x_{ij} - \beta - \gamma t_{ij})^2$$

with respect to β and γ, which yields,

$$\hat{\hat{\beta}} = x_{..} - \hat{\hat{\gamma}} t_{..},$$

$$\hat{\hat{\gamma}} = \frac{\sum_i \sum_j (x_{ij} - x_{..})(t_{ij} - t_{..})}{\sum_i \sum_j (t_{ij} - t_{..})^2},$$

$$\hat{\hat{\mu}}_{ij} = \hat{\hat{\alpha}} + \hat{\hat{\gamma}} t_{ij}. \tag{4.11}$$

Hence

$$\sum_i \sum_j (x_{ij} - \hat{\hat{\mu}}_{ij})^2 = \sum_i \sum_j (x_{ij} - x_{..})^2 - \frac{[\sum_i \sum_j (x_{ij} - x_{..})(t_{ij} - t_{..})]^2}{\sum_i \sum_j (t_{ij} - t_{..})^2}, \tag{4.12}$$

and

$$\sum_i \sum_j (\hat{\mu}_{ij} - \hat{\hat{\mu}}_{ij})^2 = \sum_i \sum_j (x_{i.} - x_{..} + \hat{\gamma}(t_{ij} - t_{i.}) - \hat{\hat{\gamma}}(t_{ij} - t_{..}))^2.$$

The statistic ω in (3.7) is given by

$$\omega = \frac{N-p-1}{p-1}\left[\frac{\sum_i \sum_j (\hat{\mu}_{ij}-\hat{\hat{\mu}}_{i})^2}{\sum_i \sum_j (x_{ij}-\hat{\mu}_{ij})^2}\right], \qquad (4.13)$$

and the uniformly most powerful invariant test of H_o rejects H_o whenever $\omega \geq c$. Under H_o, W has central F distribution with p-1, N-p-1 degrees of freedom. Hence $c = F_{1-\alpha,p-1,N-p-1}$.

From (4.12),

$$\sum_i \sum_j (x_{ij} - \hat{\hat{\mu}}_{ij})^2$$

$$= \sum_i \sum_j (x_{ij} - x_{..} - \hat{\hat{\gamma}}(t_{ij}-t_{..}))^2$$

$$= \sum_i \sum_j (x_{ij} - x_{i.} - \hat{\gamma}(t_{ij} - t_{i.}))^2$$

$$+ \sum_i \sum_j (x_{i.} - x_{..} + \hat{\gamma}(t_{ij} - t_{i.}) - \hat{\hat{\gamma}}(t_{ij} - t_{..}))^2$$

$$= \sum_i \sum_j (x_{ij} - \hat{\mu}_{ij})^2 + \sum_i \sum_j (\hat{\mu}_{ij} - \hat{\hat{\mu}}_{ij})^2.$$

Hence, for computational convenience, we can write

$$\omega = \frac{N-p-1}{p-1}\left[\frac{T_{xx} - \frac{T_{xt}^2}{T_{tt}} - (E_{xx} - \frac{E_{xt}^2}{E_{tt}})}{E_{xx} - \frac{E_{xt}^2}{E_{tt}}}\right] \qquad (4.14)$$

where

$$T_{xx} = \sum_i \sum_j (x_{ij} - x_{..})^2, \quad T_{tt} = \sum_i \sum_j (t_{ij} - t_{..})^2$$

$$T_{xt} = \sum_i \sum_j (x_{ij} - x_{..})(t_{ij} - t_{..})$$

$$E_{xx} = \sum_i \sum_j (x_{ij} - x_{i.})^2, \quad E_{tt} = \sum_i \sum_j (t_{ij} - t_{i.})^2,$$

$$E_{xt} = \sum_i \sum_j (x_{ij} - x_{i.})(t_{ij} - t_{i.}). \qquad (4.15)$$

Table 4.1. Analysis of covariance for one-way classification

Sources of variation	d.f	Sum of products xx	Sum of products xt	Sum of products tt	Adjusted sum of squares cl.f	Adjusted sum of squares sum of squares	F
Populations	p-1	P_{xx}	P_{xt}	P_{tt}	p-1	$q_1=q_o-q_2$	$\frac{N-p-1}{p-1}\,\frac{q_1}{q_2}$
Error	N-p	E_{xx}	E_{xt}	E_{tt}	N-p-1	$q_2=E_{xx}-\frac{E_{xt}^2}{E_{tt}}$	
Total	N-1	T_{xx}	T_{xt}	T_{tt}	N-2	$q_o=T_{xx}-\frac{T_{xt}^2}{T_{tt}}$	

If the hypothesis H_o is rejected, one can determine by multiple comparison methods which of the contrasts in the β_i are responsible for this. For any estimable linear parametric contrast

$$\phi = \sum_{i=1}^{p} c_i\beta_i, \text{ with } \sum_{i=1}^{p} c_i = 0,$$

its estimate, using (4.9), is given by

$$\hat{\phi} = \sum_{i=1}^{p} c_i\hat{\beta}_i = \sum_{i=1}^{p} c_i x_{i.} - \hat{\gamma}\sum_{i=1}^{p} c_i t_{i.} \tag{4.16}$$

Since

$$\text{Var}(\hat{\gamma}) = \sigma^2\Big(\sum_i\sum_j (t_{ij}-t_{i.})^2\Big)^{-1}$$

from (4.16) we get

$$\text{Var}(\hat{\phi}) = \sigma^2\Big(\sum_i \frac{c_i^2}{N_i} + \frac{(\Sigma\, c_i t_{i.})^2}{\sum_i\sum_j (t_{ij}-t_{i.})^2}\Big). \tag{4.17}$$

4.1.2. Two-way Classification (with one observation per cell): We have discussed the analysis of variance of the two-way classification with one observation per cell with two factors A and B having I and J levels respectively. We shall consider here the analysis of covariance for this classification with one concomitant variable. Let X_{ij}, $i = 1,\ldots,I$, $j = 1,\ldots,J$ be normally distributed independent random variables with

$$\mu_{ij} = E(X_{ij}) = \mu + \alpha_i + \beta_j + \gamma t_{ij}, \tag{4.18}$$

$$\operatorname{Var}(X_{ij}) = \sigma^2,$$

where μ is the grand mean, α_i is the effect of the ith level of A satisfying $\sum_{i=1}^{I} \alpha_i = 0$, β_j is the effect of the jth level of B satisfying $\sum_{j=1}^{J} \beta_j = 0$ and the t_{ij} are observations (known) on the concomitant variable. Let

$$H_\alpha : \alpha_1 = \ldots = \alpha_I = 0,$$

$$H_\beta : \beta_1 = \ldots = \beta_J = 0.$$

In the notation of Section 3.1, the parametric space π_Ω has dimension $s = I + J$ and the parametric space π_ω under the hypothesis H_α has dimension $s-r = J+1$ and under the hypothesis H_β has dimension $I + 1$. The minimum value of

$$\sum_i \sum_j (x_{ij}-\mu_{ij})^2$$

under π_Ω is obtained by minimizing

$$\sum_i \sum_j (x_{ij}-\mu-\alpha_i-\beta_j-\gamma t_{ij})^2,$$

which for fixed γ leads to

$$\mu = x_{..} - \gamma t_{..}, \quad \alpha_i = x_{i.} - x_{..} - \gamma(t_{i.} - t_{..}),$$
$$\beta_j = x_{.j} - x_{..} - \gamma(t_{.j} - t_{..}), \tag{4.19}$$

and reduces the sum of squares to

$$\sum_i \sum_j (x_{ij}-x_{i.}-x_{.j}+x_{..}+\gamma(t_{ij}-t_{i.}-t_{.j}+t_{..}))^2. \tag{4.20}$$

Minimizing (4.20) with respect to γ we get

$$\hat{\gamma} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{N_i} (x_{ij}-x_{i.}-x_{.j}+x_{..})(t_{ij}-t_{i.}-t_{.j}+t_{..})}{\sum_i \sum_j (t_{ij}-t_{i.}-t_{.j}+t_{..})^2},$$

$$\begin{aligned}
\hat{\mu} &= x_{..} - \hat{\gamma} t_{..}\,, \\
\hat{\alpha}_i &= (x_{i.} - x_{..}) - \hat{\gamma}(t_{i.} - t_{..}), \\
\hat{\beta}_j &= (x_{.j} - x_{..}) - \hat{\gamma}(t_{.j} - t_{..}), \\
\hat{\mu}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma} t_{ij}.
\end{aligned}$$

Hence

$$\sum_i \sum_j (x_{ij} - \hat{\mu}_{ij})^2$$

$$= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 - \frac{[\sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})(t_{ij} - t_{i.} - t_{.j} + t_{..})]^2}{\sum_i \sum_j (t_{ij} - t_{i.} - t_{.j} + t_{..})^2}$$

$$= E_{xx} - \frac{E_{xt}^2}{E_{tt}}\,,$$

where

$$\begin{aligned}
E_{xx} &= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2, \\
E_{xt} &= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})(t_{ij} - t_{i.} - t_{.j} + t_{..}), \\
E_{tt} &= \sum_i \sum_j (t_{ij} - t_{i.} - t_{.j} + t_{..})^2.
\end{aligned} \qquad (4.21)$$

Case a. Test of H_α. To find the test for the hypothesis H_α we have to minimize

$$\sum_i \sum_j (x_{ij} - \mu - \beta_j - \gamma t_{ij})^2$$

with respect to μ, β_j and γ, which gives

$$\hat{\hat{\mu}} = x_{..} - \hat{\hat{\gamma}}\, t_{..}\,, \quad \hat{\hat{\beta}}_j = x_{.j} - x_{..} - \hat{\hat{\gamma}}(t_{.j} - t_{..}),$$

$$\hat{\hat{\gamma}} = \frac{\sum_i \sum_j (x_{ij} - x_{.j})(t_{ij} - t_{.j})}{\sum_i \sum_j (t_{ij} - t_{.j})^2} \qquad (4.22)$$

$$\hat{\hat{\mu}}_{ij} = \hat{\hat{\mu}} + \hat{\hat{\beta}}_j + \hat{\hat{\gamma}} t_{ij}.$$

Hence from (4.21)

$$\sum_i\sum_j(x_{ij}-\hat{\hat{\mu}}_{ij})^2$$

$$=\sum_i\sum_j(x_{ij}-x_{.j})^2-\frac{[\sum_i\sum_j(x_{ij}-x_{.j})(t_{ij}-t_{.j})]^2}{\sum_i\sum_j(t_{ij}-t_{.j})^2}$$

$$=E_{xx}+A_{xx}-\frac{[E_{xt}+A_{xt}]^2}{E_{tt}+A_{tt}},$$

where

$$A_{xx}=\sum_i J(x_{i.}-x_{..})^2,\quad A_{tt}=\sum_i J(t_{i.}-t_{..})^2,$$

$$A_{xt}=\sum_i J(x_{i.}-x_{..})(t_{i.}-t_{..}). \tag{4.23}$$

Hence, from (3.7), the test statistic ω is given by

$$\omega=\frac{IJ-I-J}{(I-1)}\left[\frac{\sum_i\sum_j(x_{ij}-\hat{\hat{\mu}}_{ij})^2-\sum_i\sum_j(x_{ij}-\hat{\mu}_{ij})^2}{\sum_i\sum_j(x_{ij}-\hat{\mu}_{ij})^2}\right],$$

and the uniformly most powerful invariant test of H_α rejects H_α whenever

$$\omega\geq c, \tag{4.24}$$

where the constant c depends on the size α of the test. Since under H_α, W has central F-distribution with $(I-1, IJ-I-J)$ degrees of freedom we get $c = F_{1-\alpha,I-1,IJ-I-J}$.

Case b. Test of H_β. To find the test of H_β we have to minimize

$$\sum_i\sum_j(x_{ij}-\mu-\alpha_i-\gamma t_{ij})^2$$

with respect to μ, α_i and γ, which gives

$$\tilde{\mu}=x_{..}-\tilde{\gamma}t_{..},\quad \tilde{\alpha}_i=x_{i.}-x_{..}-\tilde{\gamma}(t_{i.}-t_{..}),$$

$$\tilde{\gamma}=\frac{\sum_i\sum_j(x_{ij}-x_{i.})(t_{ij}-t_{i.})}{\sum_i\sum_j(t_{ij}-t_{i.})^2}, \tag{4.25}$$

$$\tilde{\mu}_{ij} = \tilde{\mu} + \tilde{\alpha}_i + \tilde{\gamma} t_{ij}.$$

Hence, from (4.21),

$$\sum_i \sum_j (x_{ij} - \tilde{\mu}_{ij})^2$$

$$= \sum_i \sum_j (x_{ij} - x_{i\cdot})^2 - \frac{\left[\sum_i \sum_j (x_{ij} - x_{i\cdot})(t_{ij} - t_{i\cdot})\right]^2}{\sum_i \sum_j (t_{ij} - t_{i\cdot})^2}$$

$$= E_{xx} + B_{xx} - \frac{[E_{xt} + B_{xt}]^2}{B_{tt}}, \qquad (4.26)$$

where

$$B_{xx} = \sum_j I(X_{\cdot j} - x_{\cdot\cdot})^2, \quad B_{tt} = \sum_j I(t_{\cdot j} - t_{\cdot\cdot})^2,$$

$$B_{xt} = \sum_j I(x_{\cdot j} - x_{\cdot\cdot})(t_{\cdot j} - t_{\cdot\cdot}). \qquad (4.27)$$

Thus the uniformly most powerful invariant test of H_β rejects H_β whenever

$$\omega = \frac{IJ-I-J}{J-1} \; \frac{\sum_i \sum_j (x_{ij} - \tilde{\mu}_{ij})^2 - \sum_i \sum_j (x_{ij} - \hat{\mu}_{ij})^2}{\sum_i \sum_j (x_{ij} - \hat{\mu}_{ij})^2}$$

$$\geq c$$

where the constant c depends on the size α of the test. Since under H_β, W has a central F-distribution with (J-1, IJ-I-J) degrees of freedom we get $c = F_{1-\alpha, J-1, IJ-I-J}$.

If the hypothesis H_α is rejected, one can determine by multiple comparison methods which of the contrasts in the α_i are responsible for this rejection. The same is true for the hypothesis H_β.

4.1.3. General Linear Regression Model: Let $\underline{X} = (X_1, \ldots, X_N)'$ be normally distributed with

$$E(\underline{X}) = X\underline{\beta} + T\underline{\gamma} \qquad (4.28)$$

$$\mathrm{Cov}(\underline{X}) = \sigma^2 I$$

Table 4.2. Analysis of covariance for two-way classification

Sources of variation	Sum of products					Adjusted sum of squares	F
	d.f	xx	xt	tt	df	Sum of squares	
Between levels of A	I-1	A_{xx}	A_{xt}	A_{tt}	I-1	$q_0 = q_3 - q_2$	$\frac{IJ-I-J}{I-1}\,\frac{q_0}{q_2}$
Between levels of B	J-1	B_{xx}	B_{xt}	B_{tt}	J-1	$q_1 = q_4 - q_2$	$\frac{IJ-I-J}{(J-1)}\,\frac{q_1}{q_2}$
Error	(I-1)×(J-1)	E_{xx}	E_{xt}	E_{tt}	IJ-I-J	$q_2 = E_{xx} - \frac{E_{xt}^2}{E_{tt}}$	
Total	IJ-1	T_{xx}	T_{xt}	T_{tt}	IJ-2		
Error + levels of A	IJ-J					$q_3 = (A_{xx}+E_{xx}) - \frac{[A_{xt}+E_{xt}]^2}{A_{tt}+E_{tt}}$	
Error + levels of B	IJ-I					$q_4 = (B_{xx}+E_{xx}) - \frac{[B_{xt}+E_{xt}]^2}{B_{tt}+E_{tt}}$	

where X is the design matrix of dimension N × p, $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$, $T = (t_{ij})$ is an N × k matrix whose ith column

$$\underline{T}_i = (t_{i1}, \ldots, T_{iN})'$$

represents the N observations on the ith concomitant variable and $\underline{\gamma} = (\gamma_1, \ldots, \gamma_k)'$ is the vector of regression coefficients. The analysis of the linear model (4.28) can be done in the same way as the analysis of the linear model (3.9) in Chapter 3. No new problem arises in the adjustment for concomitant variables in the estimation of parameters and the testing of linear hypotheses. We shall give only the computational aspects of the problem.

The maximum likelihood estimate (which is the same as the least squares estimate) of $(\underline{\beta},\underline{\gamma})$ is obtained by minimizing $s^2 = (\underline{x} - X\underline{\beta} - T\underline{\gamma})'(\underline{x} - X\underline{\beta} - T\underline{\gamma})$.

Differentiating s^2 with respect to $\underline{\beta}$ and $\underline{\gamma}$ we obtain,

$$(X,T)'(X,T)\begin{pmatrix}\underline{\beta}\\ \underline{\gamma}\end{pmatrix} = (X,T)'\underline{x},$$

or equivalently,

$$\begin{aligned} X'X\underline{\beta} + X'T\underline{\gamma} &= X'\underline{x}, \\ T'X\underline{\beta} + T'T\underline{\gamma} &= T'\underline{x}. \end{aligned} \tag{4.29}$$

Let $\hat{\underline{\beta}}_o$ be the solution of

$$X'X\underline{\beta} = X'\underline{x} \tag{4.30}$$

ignoring the concomitant variables and let $\hat{\underline{\beta}}_i$ be the solution of

$$X'X\underline{\beta}_i = X'\underline{T}_i, \quad i = 1,\dots,k. \tag{4.31}$$

Define the residual sums of squares and products by

$$\begin{aligned} R_{oo} &= (\underline{x} - X\hat{\underline{\beta}}_o)'(\underline{x} - X\hat{\underline{\beta}}_o) = \underline{x}'\underline{x} - \hat{\underline{\beta}}_o X'\underline{x}, \\ R_{ii} &= (\underline{T}_i - X\hat{\underline{\beta}}_i)'(\underline{T}_i - X\hat{\underline{\beta}}_i) = \underline{T}_i'\underline{T}_i - \hat{\underline{\beta}}_i X'\underline{T}_i, \quad i = 1,\dots,k, \\ R_{oi} &= (\underline{x} - X\hat{\underline{\beta}}_o)'(\underline{T}_i - X\hat{\underline{\beta}}_i) = \underline{x}'\underline{T}_i - \hat{\underline{\beta}}\, X'\underline{T}_i, \\ &\qquad i = 1,\dots,k, \\ R_{ij} &= (\underline{T}_i - X\hat{\underline{\beta}}_i)'(\underline{T}_j - X\bar{\underline{\beta}}_j) = \underline{T}_i'\underline{T}_j - \hat{\underline{\beta}}_i\, X'\underline{T}_j, \\ &\qquad i,j = 1,\dots,k. \end{aligned} \tag{4.32}$$

From (4.29), multiplying the first equation by $\hat{\underline{\beta}}_i$ and subtracting from the ith component of the second equation we obtain

$$R_{1i}\gamma_1 + \dots + R_{ki}\gamma_k = R_{oi}, \quad i = 1,\dots,k. \tag{4.33}$$

Let $\hat{\underline{\gamma}} = (\hat{\gamma}_1,\dots,\hat{\gamma}_k)'$ be a solution of (4.33). The solution for $\underline{\beta}$ satisfying (4.29) is seen to be

$$\hat{\underline{\beta}} = \hat{\underline{\beta}}_o - \hat{\gamma}_1\hat{\underline{\beta}}_1 - \dots - \hat{\gamma}_k\hat{\underline{\beta}}_k, \tag{4.34}$$

which gives the adjustment to be made in $\hat{\underline{\beta}}_o$ due to the concomitant variables. Under π_Ω, the residual (error) sum of squares R_o^2 is given by

$$R_o^2 = \min_{\pi_\Omega} (\underline{x} - X\underline{\beta} - T\underline{\gamma})'(\underline{x} - X\underline{\beta} - T\underline{\gamma}) \tag{4.35}$$

$$= R_{oo} - \sum_{i=1}^{k} \hat{\gamma}_i R_{oi} ,$$

with N-r-k degrees of freedom, where $r = \text{rank}\ (X)$. Now to test the hypothesis $H_o: \underline{L}_1'\underline{\beta} = \theta_1, \ldots, \underline{L}_k'\underline{\beta} = \theta_k$, that is $L'\underline{\beta} = \underline{\theta}$, where $L = (\underline{L}_1, \ldots, \underline{L}_k)$ is a $p \times k$ matrix and $\underline{\theta} = (\theta_1, \ldots, \theta_k)'$ is a known vector, we have to find

$$R_2^2 = \min_{L'\underline{\beta} = \underline{\theta}} (\underline{x} - X\underline{\beta} - T\underline{\gamma})'(\underline{x} - X\underline{\beta} - T\underline{\gamma}) \tag{4.36}$$

Let R'_{oo}, R'_{oi}, R'_{ij} be the residual sums of squares and products as in (4.32) when $\underline{\beta}$ is subjected to restrictions $L'\underline{\beta} = \underline{\theta}$ and let $\hat{\gamma}'_1, \ldots, \hat{\gamma}'_k$ be the solutions of (4.33) using R'_{ij} instead of R_{ij}, then

$$R_2^2 = R'_{oo} - \sum_{i=1}^{k} \hat{\gamma}'_i R'_{oi} , \tag{4.37}$$

and the hypothesis $H_o: L'\underline{\beta} = \underline{\theta}$ is rejected whenever

$$\omega = \frac{N-r-k}{r} \; \frac{R_2^2 - R_o^2}{N-s-k} \geq F_{1-\alpha, r, N-r-k},$$

where s is the rank of (X) and r is the degrees of freedom of the hypothesis $L'\underline{\beta} = \underline{\theta}$.

We have thus far discussed the analysis of covariance of one-way and the two-way classifications and the regression model. The analysis of covariance of some specific designs will be given in Chapter 5. The analysis of covariance of higher-way classifications and other designs are straightforward extensions of the analysis of linear models (Chapter 3). In concluding this section we remark that in analysis of covariance the concomitant variates should be carefully selected and the results should be interpreted in the light of the particular concomitant variates used and the particular values of the variates obtained.

Snedecor (1934, 1946) and Cochran and Bliss (1948) presented several examples and discussed the results of several covariance analyses. Wishart (1936) obtained tests of significance in analysis of covariance. The reader is referred to Wilks (1938, 1944), De Lury (1948), Finney (1946), Fisher (1949), Federer (1963) and Nair (1942) for relevant results and discussions in covariance analysis.

4.2. EFFICIENCY, OPTIMALITY OF DESIGNS

When an experiment is laid out to estimate treatment effects with a certain number of restrictions on allotment of treatments to the experimental units it is possible to estimate how much information would have been lost by using a smaller number of restrictions. The efficiency of a particular design of experiments relative to another design of experiment with the same treatments and experimental material is defined as the efficiency of the estimate of any treatment in contrast from the particular design relative to the other design considered. In early stages of agricultural experimentation only a few experimental designs were at one's disposal and efficiency was taken as a desirable criterion for choosing the needed design for a given purpose.

Now-a-days an experimenter with modern statistical techniques at his disposal can use a variety of experimental designs for a given purpose (we shall discuss a large number of them in Chapters 5 and 6). Hence, whenever the conditions of an experiment allow the possibility of simultaneous existence of a number of experimental designs, the question of selecting an appropriate design, a design which is easy to analyse and which has some optimum properties, naturally, arises.

Our purpose here is not to state the problem of optimality of designs in the most general possible abstract setting nor to apply the discussion to the more complicated problems. Our primary aim is to stress certain ideas and principles which are most easily set forth in simple situations. We shall be concerned with verifying whether or not a given design satisfies certain optimality criteria. Our approach is in the spirit of Wald's decision theory. The reader is referred to Kiefer (1958, 1959), Kiefer and Wolfowitz (1959) for a more elegant and complete treatment of this topic.

Let $\underline{X} = (X_1,\ldots,X_p)'$ be a normally distributed N-vector with

$$E(\underline{X}) = X_d\underline{\beta}$$

$$\operatorname{Cov}(\underline{X}) = \sigma^2 I \qquad (4.38)$$

where $\underline{\beta} = (\beta_1,\ldots,\beta_p)'$, with unknown components β_i, X_d is a known $N \times p$ matrix which depends on the design d and which is subject to choice by the experimenter. Denote by Δ the set of all experimental designs available to the experimenter and by Δ_R (the class of all randomized designs), the class of all probability measures on Δ. A randomized design δ is used by choosing a d from Δ according to this measure δ and then using the selected design.

We shall call $A_d = (X_d)'(X_d)$, the information matrix of the design d.

Let

$$\phi_i = \underline{L}_i'\underline{\beta}, \; i = 1,\ldots,s,$$

where $\underline{L}_i$ is a p-vector of known constants. Suppose that we are interested in statistical inferences concerning ϕ_j, $j = 1,\ldots,s$. Let Δ' be the class of designs in which the ϕ_j are estimable and let $\sigma^2 V_d$ be the covariance matrix of the best linear estimator of the ϕ_j for a d in Δ'. For testing the hypothesis

$$H_o\colon \phi_1 = \cdots = \phi_s = 0,$$

using a design d in Δ', let $\beta_\psi(\delta^2)$ be the infimum of the power function of the test ψ over all alternatives

$$\sum_{j=1}^{s} \frac{\phi_j^2}{\sigma^2} = \delta^2$$

and let $\overline{\beta}(d,\delta^2,\alpha)$ be the supremum of $\beta_\psi(\delta^2)$ over all tests ψ of size α.

Definition 4.1. (M-optimum). For, $\delta^2 > 0$, $0 < \alpha < 1$, a design d* is said to be M-optimum in Δ if

$$\bar{\beta}(d^*, \delta^2, \alpha) = \sup_{d \varepsilon \Delta} \bar{\beta}(d, \delta^2, \alpha) . \tag{4.39}$$

Definition 4.2. (L-optimum). A design d* is said to be L-optimum in Δ if

$$\lim_{\delta^2 \to 0} \left[\frac{\bar{\beta}(d^*, \delta^2, \alpha) - \alpha}{\sup\limits_{d \varepsilon \Delta} (\bar{\beta}(d, \delta^2, \alpha) - \alpha)} \right] = 1 \tag{4.40}$$

Definition 4.3. (D-optimum). A design d* is said to be D-optimum in Δ if $d^* \varepsilon \Delta'$ and if

$$|V_{d^*}| = \operatorname*{Inf}_{d \varepsilon \Delta'} |V_d| . \tag{4.41}$$

Definition 4.4. (E-optimum). A design d* is said to be E-optimum in Δ if $d^* \varepsilon \Delta'$ and if the maximum characteristic root of V_{d^*} is less than or equal to the maximum characteristic root of V_d for any $d \varepsilon \Delta'$.

Definition 4.5. (A-optimum). A design d* is said to be A-optimum in Δ if $d^* \varepsilon \Delta'$ and if

$$\operatorname{tr} V_{d^*} = \min_{d \varepsilon \Delta'} \operatorname{tr} V_d .$$

It may be noted that these definitions do not depend on whether σ^2 is known or not.

M-optimality, the strongest and the least artificial among the above five criteria is extremely difficult to characterize even in very simple situations. To understand the implication of L-optimality let us consider the following example.

Example 4.1. Consider a one-way analysis of variance set-up with linear model as given in (4.38). Suppose that with rank of $X_d = 2$, $N = 4$, $\sigma^2 = 1$ and with the problem of testing $H_0: \beta_1 = \beta_2 = 0$, we use the non-randomized design d with two observations from each population. Suppose also that we are interested to find a test of size α for H_0 which maximizes the minimum power on each contour $\beta_1^2 + \beta_2^2 = \delta^2$. Hunt and Stein in their famous unpublished work (see Kiefer, 1958) showed that the usual test based on chisquare with 2 degrees of freedom does the job.

Let r_d be the rank of X_d. Then, from the results presented in Chapter 3, there exists an $r_d \times p$ matrix L_d of rank r_d such that the distribution of $\underline{X}$ depends on $\underline{\beta}$ only through $L_d\underline{\beta}$ whose best linear estimator is given by $L_d\hat{\underline{\beta}}_d$ where $\hat{\underline{\beta}}_d$ is any solution of $A_d\underline{\beta} = X'_d\underline{x}$. If $r_d = p$ then L_d is the identity matrix.

Now suppose that we first start with a design d, but due to missing observations we end up with the design matrix d' such that $A_{d'} = (X_{d'})'(X_{d'}) = \begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$.

Now, presumably, our equal interest (with respect to β_1, β_2) in the alternatives $\beta_1^2+\beta_2^2 = \sigma^2$ should not be conveniently changed in the way to justify the use of χ^2 test based on $\hat{\beta}_{d_1}^2 + 2\hat{\beta}_{d_2}^2$, by the same accident which reduces us from d to d'. But if we want a size α test which maximizes the minimum power on the contour $\beta_1^2 + \beta_2^2 = \delta^2$ then this test no longer satisfies this criterion and should not be used.

In the above example if we let δ^2 tend to zero and rephrase our criterion in terms of the derivatives of the power function the problem reduces to that of finding a regular Neyman and Pearson's type c test (Neyman and Pearson, 1938). In the case of the design d, the appropriate test is then the usual one based on the χ^2-distribution while in the case of design d' rejection region is of the form $a\,\hat{\beta}_{d_1}^2 + 2b\hat{\beta}_{d_2}^2 > 1$ where the positive constants a, b can be obtained numerically (see Kiefer, 1959). Thus for non-randomized designs if we want to find that design and the associated test which approximately maximize the minimum power on small spheres about the null hypothesis ($\delta^2=0$) we can conclude that a regular type c region is to be used with whatever design is chosen. This is called the L-optimum design. Obviously, thus, L-optimality is a local version of M-optimality.

E-optimality was first used by Wald (1943) and Ehrenfeld (1955) in testing hypotheses. The analysis of variance F-test of size α is E-optimum for all δ^2. For point estimation an E-optimum design minimizes the maximum overall $(a_1,\ldots,a_p)$ with $\sum_{i=1}^{s} a_i^2=1$ of the variance of the best linear estimator of $\sum_{i=1}^{s} a_i\phi_i$.

An A-optimum design maximizes the average variance of the best linear estimator of $\phi_1,\ldots,\phi_s$ and thus of any s linear parametric functions obtained from the ϕ_i by orthogonal transformation. A D-optimum design minimizes the generalized variance of the best linear estimators of the ϕ_i and thus when σ^2 is known, it minimizes the volume of the smallest confidence region on $\phi_1,\ldots,\phi_s$ for any confidence coefficient. For testing of hypotheses a D-optimum design achieves a test whose power function has maximum Gaussian curvature at the null hypothesis, among all locally unbiased tests of a given size α.

Because of the ad hoc assumption that the F-test, or the corresponding χ^2 test for σ^2 known, should always be used, D-optimality and L-optimality although local properties seem more reasonable than E-optimality.

D-optimality has an appealing invariant property which is not possessed by any other criterion. Let $\phi'_1,\ldots,\phi'_s$ be related to $\phi_1,\ldots,\phi_s$ by a nonsingular linear transformation. Then if d^* is D-optimum for $\phi_1,\ldots,\phi_s$ then it is also D-optimum for $\phi'_1,\ldots,\phi'_s$. For testing of hypotheses D-optimality is also invariant under the group of nonsingular transformations. Kiefer (1958), by means of examples has established that E-optimality is the least satisfactory of the criteria considered here. He has also proved that if r_d = rank of X_d, is constant for $d \in \Delta'$ and if d^* is D-optimum and V_{d^*} is a multiple of the identity matrix, then d^* is E-optimum and L-optimum.

For some other optimality criteria in testing of hypotheses the reader is referred to Giri and Kiefer (1964).

EXERCISE 4

1. The following data relating to the first lactation yield in 10 lbs units of dams and daughters of 4-sires are taken from the records of Kankrej herd maintained at the Institute of Agriculture, Anand, India (Reproduced from the National Index of field

experiments M.P.6, Institute of Agricultural Research Statistics with the kind permission of Prof. M.N.Das).

Table 4.3. Lactation yield of dams and daughters for 4 different sires

Sire 1		Sire 2	
Dam's yield	Daughter's yield	Dam's yield	Daughter's yield
216	258	153	112
216	178	276	269
269	243	260	247
236	178	299	198
153	111	298	357
280	233	266	327
190	60	281	200
340	262	186	190
216	191	188	261
277	138	168	235
121	197	251	165
261	234		
190	123		
Sire 3		Sire 4	
274	313	340	340
375	374	251	247
312	320	169	193
229	532	208	235
171	203	182	252
197	269	221	280
226	153	193	194
260	228	312	169
166	184	299	242
252	383	188	253

Analyse the data using Dam's yields as concomitant variables to test the hypothesis of equality of daughters' mean lactation yields for different sires.

2. The following data (Wishart, 1938) relate to the initial weight and growth rate of 30 pigs classified according to pen, sex and type of food given.

Table 4.4. Initial weight and growth rate in lb per week for 30 different pigs

Pen	Food	Sex	Initial weight (in lbs)	Growth rate per week (in lbs)
I	A	G	48	9.94
	B	G	48	10.00
	C	G	48	9.75
	C	H	48	9.11
	B	H	39	8.51
	A	H	38	9.52
II	B	G	32	9.24
	C	G	28	8.66
	A	G	32	9.48
	C	H	37	8.50
	A	H	35	8.21
	B	H	38	9.95
III	C	G	33	7.63
	A	G	35	9.32
	B	G	41	9.34
	B	H	46	8.43
	C	H	42	8.90
	A	H	41	9.32
IV	C	G	50	10.37
	A	H	48	10.56
	B	G	46	9.68
	A	G	46	10.98
	B	H	40	8.86
	C	H	42	9.51
V	B	G	37	9.67
	A	G	32	8.82
	C	G	30	8.57
	B	H	40	9.20
	C	H	40	8.76
	A	H	43	10.42

Analyse the data using the initial weights as concomitant variables to test the hypothesis of equality of mean food effects on pigs.

3. In a simple linear regression of Y on X, suppose n points of X have to be selected (without error) over [a,b] and then values of Y have to be observed corresponding to each selected x-point in order to get efficient estimate of the regression coefficient β in the linear model

$$E(Y|X = x) = \alpha + \beta(x-\bar{x}),\ \mathrm{Var}(Y|X = x) = \sigma^2,$$

where $\bar{x}$ represents the mean of the x-points selected. Prove that an efficient design in the sense of minimizing Var(b), where b is the usual least-square estimate of β is provided by the following: If n = 2m, choose m of the x's equal to a and in others equal to b; if n = 2m+1, choose 2m x's as above and one more equal to (a+b)/2. [Note that at least two of the x's must be different from each other.]

REFERENCES

Cochran, W.G. and Bliss, C.I. (1948). Discriminant functions with covariance. Ann. Math. Statist. 19, 151-176.

De Lury, D.B. (1948). The analysis of covariance. Biometrica, 4, 153-170.

Enrenfeld, S. (1955). On the efficiency of experimental designs. Ann. Math. Statist. 26, 247-255.

Federer, W.T. (1963). Experimental designs, theory and applications. The Macmillan Co., New York.

Finney, D.J. (1946). Standard errors of yields adjusted for regression on an independent measurement. Biometrics, 2,53-55.

Fisher, R.A. (1949). The design of experiments. Oliver and Boyd, London.

Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax properties of multivariate tests. Ann.Math.Statist.35, 21-35.

Kiefer, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. Ann.Math. Statist. 29, 675-699.

Kiefer, J. (1959). Optimum experimental designs. J. R. Stat. Soc. B. 21, 272-319.

Kiefer, J. and Wolfowitz, J. (1959). Optimum design in regression problems. Ann.Math.Statist.30.

Nair, K.R. (1942). Efficiency of the adjustment for concomitant characters in biological experiments. Sankhya, 6, 167-174.

Neyman, J. and Pearson, E.S. (1938). Contributions to the theory of testing statistical hypotheses, III, Stat. Res. Memoirs, 2, 25-27.

Snedecor, G.W. (1934). Calculation and interpretation of analysis of variance and covariance. Collegiate Press Inc. Ames.

Snedecor, G.W. (1946). Statistical methods. Iowa State College Press. Ames.

Wald, A. (1943). On the efficient design of statistical investigations. Ann.Math.Statist.14, 134-140.

Wilks, S.S. (1938). The analysis of variance and covariance in nonorthogonal data. Metron, 13, 141-154.

Wilks, S.S. (1944). Mathematical statistics. Princeton Univ. Press.

Wishart, J. (1938). Growth rate determinations in nutrition studies with the bacon pig and their analysis. Biometrika, 30, 16-28.

Wishart, J. (1936). Tests of significance in analysis of covariance. J.R.S.S. Suppl. 3, 79-82.

CHAPTER 5. EXPERIMENTAL DESIGNS AND THEIR ANALYSIS

Various types of experimental designs are in use in experimental sciences and the analysis of these designs involves the linear models of general linear hypothesis (Chapters 3 and 4). First we shall consider some basic complete block designs, then we shall consider the case of incomplete block designs.

The main aim of a scientific experiment is to know something which is not known about a group of treatments and to prove or disprove certain hypotheses about the treatment effects. The term "treatment" is a general one and is used for a group, a category or a class. The term experimental unit will refer to the type of experimental material that is being used to receive the application of various treatments under consideration. For example, if one is interested in investigating the effect of various fertilizers on potted plants, the potted plant will be the experimental unit. In agricultural experiments one uses a plot of la as an experimental unit. Once a sufficient number of experimental units has been selected, it is then necessary to allocate the treatments to the experimental units in a random fashion. This random allocation is necessary to protect against any bias in the experiment resulting from the influence of some extraneous unknown factors affecting the experiment. For example in agricultural experiments, if the experimental units (plots) are not homogeneous with regard to soil fertility, we are sure to get different responses for treatments, even when they are alike, if specific treatments are allocated to experimental units with specific fertility characteristics. If there is some amount of replication in the experiment, random allocation also enables one to attach valid probabilistic statements to estimated treatment differences. Fisher (1949) has stated that randomization and replication are necessary to obtain a valid estimate of error variance.

Design of experiments is the method by which treatments are placed at random on the experimental units in such a way that the responses are estimated with the utmost precision possible. If the randomization process is such that every experimental unit has an equal chance of receiving each treatment, we call it a complete randomization. However, due to the nature of the experiment, restrictions on randomization may be inherent in the design to facilitate a gain in precision.

Most experimental designs used in agricultural field experiments are block designs. The available experimental units (plots) are grouped in blocks of the same size (that is, the same number of experimental units), having more or less identical fertility characteristics. The fertility characteristics of different blocks are, however, usually different. If the size of the block is equal to the number of treatments and there is a random allocation of each treatment in each block, we call the design a complete block design. However, ~~if the~~ number of treatments is so large that a full replication in each block makes it too heterogeneous in fertility, it may be advisable to use smaller, more homogeneous blocks not containing a full replicate. Experimental designs with blocks containing an incomplete replication of treatments are called incomplete block designs. In the category of complete block designs we shall consider randomized block and latin square designs, and in the category of incomplete block designs we shall consider lattice, balanced incomplete block and partially balanced incomplete block designs.

5.1. COMPLETELY RANDOMIZED DESIGN

The simplest of all designs having a random arrangement is the completely randomized design. It is defined to be one when a group of ν treatments are randomly allocated to the whole set of experimental units, without making any effort to group the latter in any way for more homogeneity.

The number of replications for different treatments need not be equal, it may vary from treatment to treatment depending on the knowledge (if any) on the variability of the observations on individual treatments and on the accuracy required for the estimate of the individual treatment effects. In this design the entire experimental area or material is divided into a number of experimental units, say N. Given the total number of experimental units available for the experiment, the number of replications for different treatments are fixed in advance. Let us number the treatments from 1 to ν and let N_i be the number of replications required for the ith treatment such that $\sum_{i=1}^{\nu} N_i = N$. A random selection of N_1 experimental units out of the N available is made and the treatment number 1 is applied to these units. A random selection of N_2 of the remaining $N-N_1$ units is made and number 2 treatment is applied to these N_2 units. This procedure continues until all treatments have been applied. Unless practical limitation dictates otherwise or unless some treatments are more variable than others or are of greater interest an equal number of replications on each treatment should be made.

The completely randomized design is useful in laboratory techniques, in green house studies, in experiments with animals, etc. It is usually very well suited only for a small number of treatments and for homogeneous experimental materials. This design is seldom used for field experiments.

5.1.1. Analysis of Completely Randomized Design: Let x_{ij}, $j=1,\ldots,N_i$, $i = 1,\ldots,\nu$ denote the individual measurement of the jth experimental unit for the ith treatment and let

$$x_{i.} = \frac{1}{N_i} \sum_j x_{ij} \, , \quad x_{..} = \frac{1}{N} \sum_i \sum_j x_{ij} .$$

The corresponding random variables X_{ij} will be assumed to be normally distributed independent random variables with

$$\begin{aligned} E(X_{ij}) &= \mu + \alpha_i \\ \operatorname{Var}(X_{ij}) &= \sigma^2 \end{aligned} \qquad (5.1)$$

where μ is the overall mean and α_i is the ith treatment effect satisfying $\sum_i \alpha_i = 0$. In what follows in this chapter we shall use the term measurement in the general sense indicating that it may be a count, or a grade or a measurement of a certain characteristic dependent on treatments.

We shall use "S.S." to denote sum of squares and "d.f." to denote degrees of freedom. The model (5.1) corresponds to the linear model of the one-way classification of Section 3 and the ANOVA table for testing $H_o:\alpha_1 = \dots = \alpha_\nu = 0$ is given by

Table 5.1. ANOVA for completely randomized design

Sources of variation	d.f.	S.S.	Expected mean S.S.	F
Between treatments	$\nu-1$	$q_1 = \sum_i N_i(x_{i.}-x_{..})^2$	$\sigma^2 + \frac{1}{\nu-1}\sum_i N_i\alpha_i^2$	$\frac{N-\nu}{\nu-1}\frac{q_1}{q_2}$
Error	$N-\nu$	$q_2 = \sum_i \sum_j (x_{ij}-x_{i.})^2$	σ^2	
Total	$N-1$	$\sum_i \sum_j (x_{ij}-x_{..})^2$		

From Section 3.5 the uniformly most powerful invariant test of H_0 (of level α) rejects H_0 whenever $\frac{N-\nu}{\nu-1}(q_1/q_2) \geq F_{\alpha,\nu-1,N-\nu}$. If, in practice, H_0 is rejected at a given level of significance α, one has to use the multiple comparison methods of Chapter 3 to divide the entire set of treatments in different groups such that the treatments in the same group have the same treatment effects but the treatments in different groups have different treatment effects. The analysis of covariance of this design is the same as that of the one-way classification discussed in Section 4.1.1.

5.2. RANDOMIZED BLOCK DESIGN

The main disadvantage of completely randomized design is that it is usually suited only for a small number of treatments and for homogeneous experimental material. In any experiment where a large number of treatments are used, a relatively large number of experimental units are required. This generally increases the variation among treatment responses and the completely randomized design is not very helpful. If the whole of the experimental area or material is not homogeneous, it may be possible to group the material into blocks of ν units each such that the units within a block resemble one another more closely than units in different blocks. In a randomized block design the randomization is restricted and treatments are randomly allocated within each block. If there are b such blocks we have two-way layouts, the factors being treatments at ν levels and blocks at b levels. Thus a two-way layout is called a randomized block design if within each block the ν-treatments are assigned at random to the ν experimental units in such a way that each of the $\nu!$ ways of assigning the treatments to the units has the same probability of being adopted in the experiment and the assignments in different blocks are statistically independent. (See Cochran and Cox (1951) for a table of random permutation.) This design was originally developed by Fisher (1926) and has become popular in a large number of field experiments.

5.2.1. Analysis of Randomized Block Design: Let x_{ij}, $j=1,\ldots,\nu$, $i = 1,\ldots,b$ denote the individual measurement of the jth treatment in the ith block. The corresponding random variable X_{ij} will be assumed to be independently, normally distributed with

$$E(X_{ij}) = \mu + b_i + t_j \tag{5.2}$$

$$\mathrm{Var}(X_{ij}) = \sigma^2$$

where μ is the overall mean, t_j is the jth treatment effect satisfying $\sum_j t_j = 0$ and b_i is the ith block effect satisfying $\sum_i b_i = 0$. The hypotheses of interest are

$$H_T: t_1 = \ldots = t_\nu = 0,$$

$$H_B: b_1 = \ldots = b_b = 0 \,. \tag{5.3}$$

In the contest of randomized block design the most meaningful hypothesis is H_T. However if H_B is accepted then it gives evidence that blocking is not necessary for future experimentation. Identifying this with the two-way classification with one observation per cell the ANOVA table for testing H_T and H_B is in Table 5.2

Table 5.2. ANOVA for randomized block design

Source of variation	d.f.	S.S.	Expected mean S. S.	F
Blocks	$b-1$	$q_1=\sum_i \nu(x_{i.}-x_{..})^2$	$\sigma^2+\frac{1}{b-1}\sum_i \nu b_i^2$	$(\nu-1)\frac{q_1}{q_3}$
Treatments	$\nu-1$	$q_2=\sum_j b(x_{.j}-x_{..})^2$	$\sigma^2+\frac{1}{\nu-1}\sum_j bt_j^2$	$(b-1)\frac{q_2}{q_3}$
Error	$(b-1)(\nu-1)$	$q_3=\sum_i\sum_j(x_{ij}-x_{i.}-x_{.j}+x_{..})^2$	σ^2	
Total	$b\nu-1$	$\sum_{ij}(x_{ij}-x_{..})^2$		

The hypothesis H_B of equality of block effects is rejected at the level of significance α if

$$(v-1)\frac{q_1}{q_3} \geq F_{1-\alpha,b-1,(v-1)(b-1)} \tag{5.4}$$

and the hypothesis H_T of equality of treatment effects is rejected at the level of significance α if

$$(b-1)\frac{q_2}{q_3} \geq F_{1-\alpha,v-1,(v-1)(b-1)} \tag{5.5}$$

If, in practice, H_T is rejected at a given level of significance α, one has to use the multiple comparison methods of Chapter 4 to divide the entire set of treatments into different subgroups such that the treatments of the same subgroup have the same treatment effect and those in different subgroups have different treatment effects. From Section 3.7 it follows that these two tests are best invariant for the corresponding problems.

5.2.2. Missing Observations: Allan and Wishart (1930) were the first to present formulae for estimating one missing observation in the randomized block design and in the Latin square design (to be discussed in 5.3). Yates (1933) showed that their formulae resulted in minimizing the error sum of squares in the ANOVA table for the randomized block design with a missing observation. He then presented an iterative method of estimating several missing observations in randomized block designs. He also showed that the ANOVA technique for testing linear hypotheses in randomized block design with missing observations is still valid provided that the number of missing observations is not large and that the number of degree of freedom for the error sum of squares for the ordinary randomized block design is reduced by the number of missing observations. In case a treatment or a block is completely missing it makes no sense to estimate it and the analysis should be carried out omitting entirely the affected treatment or block.

Let us assume that the observation x_{ij} of jth treatment in the ith block is missing and let

$$T'_{i.} = \sum_{\substack{k \\ k\neq j}} x_{ik}, \quad T'_{.j} = \sum_{\substack{\ell \\ \ell\neq i}} x_{\ell j}, \quad T'_{..} = \sum_{\ell} \sum_{\substack{k \\ \ell\neq i,\, k\neq j}} x_{\ell k} \,.$$

From Table 5.2, the terms containing x_{ij} in the error sum of squares of the ordinary randomized block design can be written as

$$x_{ij}^2 + \frac{(T'_{..}+x_{ij})^2}{bv} - \frac{(T'_{i.}+x_{ij})^2}{v} - \frac{(T'_{.j}+x_{ij})^2}{b} \qquad (5.6)$$

Thus for variation in x_{ij} the error sum of squares is minimum when

$$2x_{ij} + 2\frac{(T'_{..}+x_{ij})}{bv} - 2\frac{(T'_{i.}+x_{ij})}{v} - 2\frac{(T_{.j}+x_{ij})}{b} = 0$$

giving

$$\hat{x}_{ij} = \frac{bT'_{i.} + vT'_{.j} - T'_{..}}{(b-1)(v-1)} \qquad (5.7)$$

This value of x_{ij} is taken for the missing value and the approximate ANOVA table is completed in the ordinary fashion except that the number of degrees of freedom for the error sum of squares is reduced by 1. The test of H_T derived from this analysis of variance is known to be biased in the sense that the expected value of treatment mean square is greater than the expected value of the error mean square under H_T. However if this test, as computed from this approximate analysis of variance table accepts there is no need to compute the exact test of H_T which is obtained by replacing in the approximate F-statistic the approximate treatment sum of squares by the following accurate treatment sum of squares, the accurate treatment sum of squares

$$= \left(\sum_{\substack{\ell=1\\ \ell\neq i,}}^{b} \sum_{\substack{k=1\\ k\neq j}}^{v} x_{\ell k}^2 - \frac{T'^2_{..}}{bv-1} \right) - \text{(error sum of squares calculated from the approximate analysis of variance table)}$$

$$-\left(\frac{T_{1.}^2}{v} + \dots + \frac{T_{i-1}^2}{v} + \frac{T_i'^2}{v-1} + \frac{T_{i+1}^2}{v} + \dots + \frac{T_{b.}^2}{v} \right)$$

A simple method when there are several missing observations is to replace as a first approximation each missing observation by the mean of the block in which it lies. The formula (5.7) is then used to get the second approximation in order, utilizing either the block mean or the new approximation if it has already been obtained for all missing values except the one being estimated. Such a procedure converge rapidly in general. If x_{ij} is missing, the estimate $\hat{t}_j$ of t_j is given by

$$\begin{aligned} \hat{t}_j &= \frac{1}{b}\,(T'_{.j} + \hat{x}_{ij}) \\ &= \frac{1}{b}\,\left[T'_{.j} + \frac{vT'_{.j} + bT'_{i.} - T'_{..}}{(b-1)(v-1)}\right] \end{aligned} \tag{5.8}$$

Thus the variance of the estimator $\hat{t}_j$ corresponding to the estimate $\hat{t}_j$ is given by

$$\begin{aligned} \mathrm{Var}(\hat{t}_j) &= \frac{\sigma^2}{(b-1)}\,\left[1 + \frac{1}{b(\nu-1)}\right] \\ &= \frac{\sigma^2}{b}\,\left[1 + \frac{v}{(b-1)(\nu-1)}\right] \end{aligned} \tag{5.9}$$

Obviously the variance of the estimator of any other treatment effect t_k without missing observations is given by

$$\mathrm{Var}(\hat{t}_k) = \frac{\sigma^2}{b}\,.$$

Furthermore for any treatment contrast $\sum_j a_j\hat{t}_j$ with $\sum_j a_j = 0$, where the ith treatment has one missing observation the variance of the estimator $\sum_j a_j\hat{t}_j$ is given by

$$\mathrm{Var}\Big(\sum_j a_j\hat{t}_j\Big) = \frac{\sigma^2}{b} \sum_{j\neq i} a_j^2 + a_i^2\,\frac{\sigma^2}{b}\Big(1+\frac{v}{(b-1)(\nu-1)}\Big) \qquad (5.10)$$

5.1. A Worked out Example. Consider Problem 19 of Exercise 3. The following diagram illustrates the experimental layout of this two-way classification with 9 treatments numbered 1,2,...,9. The treatments are randomly assigned to experimental units within blocks. The corresponding yields of rabi crop are given below the treatment number.

Block 1

5	1	9	2	4	6	7	8	10	3
5.50	5.62	5.12	5.00	4.00	3.50	5.50	6.12	5.00	4.75

Block 2

10	5	8	7	9	3	1	6	2	4
5.37	4.75	4.75	4.50	5.75	4.75	4.50	5.50	5.50	4.25

Block 3

2	7	1	4	8	9	3	5	10	6
4.00	6.00	3.87	5.00	5.00	5.25	4.50	4.50	5.75	4.50

Block 4

1	4	9	8	10	2	6	3	7	5
3.50	4.00	3.75	4.75	5.50	3.50	3.70	5.00	5.00	4.50

Since

$$F_{1-\alpha,3,27} = \begin{cases} 3.65, & \text{if } \alpha = .025, \\ 2.98, & \text{if } \alpha = .05, \\ 2.17, & \text{if } \alpha = .10, \end{cases}$$

from (5.5) we accept the hypothesis of the equality of treatment effects at least for $\alpha \leq .10$.

Table 5.3. Analysis of Variance

Sources of variation	d.f.	S.S.	Mean S.S.	F
Blocks	3	3.0131	1.0044	
Treatments	9	5.9305	0.6589	1.62
Error	27	10.9553	.4057	
Total	39	19.8989		

5.2.3. Analysis of Covariance: The analysis of covariance of the randomized block design with one concomitant variable is that of a two-way classification with one observation per cell, discussed in 4.1, the levels of the factor A being blocks (b=I) and the levels of the factor B being treatments (v = J).

5.2.4. Randomization Model of the Randomized Block Design

(a) Estimation. In the linear model (5.2) for the analysis of the randomized block design we have assumed that the errors in the observations x_{ij} are independently, normally distributed with the same variance σ^2. Here we shall deal with a linear model (called the randomization model) which does not make any such assumptions about errors, but takes into account the actual randomization employed in assigning the treatments to the experimental units. We first formulate such a randomization model for the randomized block design and then discuss the problem of estimation and testing with such a model. Randomization models were first formulated by Neyman (1923) for the completely randomized block design, and later (Neyman, (1933)), for the randomized block design.

Consider the set up of a randomized block design where ν treatments are compared on b experimental plots and that νb plots are grouped into b blocks of ν plots each such that in each block ν treatments are assigned to the ν plots independently for the b blocks so that all $(\nu!)^b$ assignments of treatments to units have the same probability of being used in the experiment.

Let x_{ijk}, μ_{ijk} denote the observed and the true measurements with treatment j on plot k in the ith block. These conceptual quantities are regarded as the observed and the expected values of the random measurements X_{ijk} satisfying

$$E(X_{ijk}) = \mu_{ijk} \, , \tag{5.11}$$

$j,k = 1,\ldots, v$, $i = 1,\ldots,b$. Let

$$x_{ijk} = \mu_{ijk} + e_{ijk} \tag{5.12}$$

The e_{ijk} are called technical errors of measurements and are regarded as random variables with

$$E(e_{ijk}) = 0 \, . \tag{5.13}$$

Using the usual dot notations for means (Chapter 3) we can write

$$\mu_{ijk} = \mu + \alpha_i + t_j + \gamma_{ij} + \eta_{ijk} \, ,$$

$$x_{ijk} = \mu + \alpha_i + t_j + \gamma_{ij} + \eta_{ijk} + e_{ijk} \tag{5.14}$$

where

$$\mu = \mu_{...} \, , \quad \alpha_i = \mu_{i..} - \mu_{...} \, , \quad t_j = \mu_{.j.} - \mu_{...} \, ,$$

$$\gamma_{ij} = \mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...} \, ,$$

$$\eta_{ijk} = \mu_{ijk} - \mu_{ij.} \, .$$

The quantities η_{ijk} are called unit errors (unknown constants).

Obviously we do not observe the conceptual quantities x_{ijk}. Merely we observe the measurement x_{ij} of the treatment j in the ith block. From (5.14) we can write

$$x_{ij} = \mu + \alpha_i + t_j + \gamma_{ij} + \eta_{ij} + e_{ij}$$

with

$$\sum_i \alpha_i = 0, \quad \sum_j t_j = 0, \quad \sum_j \gamma_{ij} = \sum_i \gamma_{ij} = 0, \tag{5.15}$$

$$\eta_{ij} = \sum_k \sigma_{ijk} \eta_{ijk} \, , \quad e_{ij} = \sum_k \sigma_{ijk} e_{ijk} \tag{5.16}$$

where v^2b random variables δ_{ijk} are defined by

$$\sigma_{ijk} = \begin{cases} 1, & \text{if the jth treatment occurs on plot k in the ith block,} \\ 0, & \text{otherwise.} \end{cases} \tag{5.17}$$

The joint distribution of δ_{ijk} is completely determined by the randomization described above and is independent of the distribution of the e_{ijk}.

Lemma 5.1.

(i) $E(\delta_{ijk}) = \frac{1}{\nu}$,

(ii) $E(\delta_{ijk}\ \delta_{i'j'k'}) = \frac{1}{\nu^2}$, if $i \neq i'$,

(iii) $$E(\delta_{ijk}\ \delta_{ij'k'}) = \begin{cases} \nu^{-1}\ \delta_{kk'} , & \text{if } j = j' , \\ \nu^{-1}\ (\nu-1)^{-1}(1-\delta_{kk'}), & \text{if } j \neq j' , \end{cases}$$

where $\delta_{kk'} = 1$, if $k = k'$ and is zero otherwise.

Proof.

(i) $E(\delta_{ijk}) = P\{\delta_{ijk} = 1\} = \frac{1}{\nu}$.

(ii) Since for $i \neq i'$ the randomizations in the ith and the i'th blocks are independent,

$$\begin{aligned} E(\delta_{ijk}\ \delta_{i'j'k'}) &= P\{\delta_{ijk}\ \delta_{i'j'k'} = 1\} \\ &= P\{\delta_{ijk} = 1,\ \delta_{i'j'k'} = 1\} \\ &= \frac{1}{\nu^2} \quad \text{if } (i \neq i'). \end{aligned}$$

(iii) $E(\delta_{ijk}\ \delta_{ij'k'}) = P\{\delta_{ijk}\delta_{ij'k'} = 1\}$

$$= P\{\delta_{ij'k'} = 1 | \delta_{ijk} = 1\}\ P\{\delta_{ijk} = 1\}$$

$$= \begin{cases} \nu^{-1}\ \delta_{kk'} , & \text{if } j = j' , \\ \nu^{-1}(\nu-1)^{-1}(1-\delta_{kk'}), & \text{if } j \neq j' . \end{cases}$$ Q.E.D.

Since $\sum_k \eta_{ijk} = 0$, from (5.16) we get

$$E(\eta_{ij}) = 0 \tag{5.18}$$

Since the σ_{ijk} and the e_{ijk} are independent and $E(e_{ijk}) = 0$, from (5.16) we get

$$E(e_{ij}) = \sum_k E(\sigma_{ijk})E(e_{ijk}) = 0 . \qquad (5.19)$$

For any contrast $\phi = \sum_j c_j t_j$ with $\sum_j c_j = 0$ an unbiased estimate of ϕ is

$$\hat{\phi} = \sum_j c_j x_{.j} ,$$

as, from (5.15), (5.18) and (5.19), we get

$$E(X_{.j}) = \mu + t_j .$$

Though theoretically the $\mathrm{Var}(\hat{\phi})$ can be computed from the moments of δ_{ijk} as given in Lemma 5.1, it is complicated and involves the η_{ijk} and the variances and covariances of the e_{ijk}. Some simplification is introduced by assuming complete additivity in the linear model which we will discuss in connection with the efficiency of the randomized block design relative to the completely randomized design. It is not possible to find an unbiased estimate of $\mathrm{Var}(\hat{\phi})$ without introducing some such simplifying assumption.

An overestimate of $\mathrm{Var}(\hat{\phi})$ can be obtained by first estimating ϕ separately in each block by

$$\hat{\phi}_i = \sum_j c_j x_{ij} , \qquad (5.20)$$

$i = 1,\ldots,b$. Then finding

$$s^2 = \frac{1}{b-1} \sum_i (\hat{\phi}_i - \hat{\phi}_.)^2 \qquad (5.21)$$

where $\hat{\phi}_. = b^{-1} \sum \hat{\phi}_i$.

Lemma 5.2.

$$E\left(\frac{s^2}{b}\right) = b^{-1}(b-1)^{-1} \sum_i \lambda_i^2 + \mathrm{Var}(\hat{\phi}) , \qquad (5.22)$$

where

$$\lambda_i = \sum_j c_j \gamma_{ij}, \ i = 1,\ldots,b.$$

Proof. From (5.15), with $f_i = \sum_j c_j(\eta_{ij} + e_{ij})$,

$$\hat{\phi}_i = \phi + \lambda_i + f_i \ . \qquad (5.23)$$

From (5.18) – (5.19) we get $E(f_i) = 0$. Furthermore, since the f_i are associated with different blocks and the randomizations in different blocks are mutually independent, from (5.23) with $f_. = \frac{1}{b} \sum_i f_i$

$$\sum_i (\hat{\phi}_i - \hat{\phi}_.)^2 = \sum_i (\lambda_i + f_i - f_.)^2$$

$$= \sum_i \lambda_i^2 + \sum_i (f_i - f_.)^2 + 2 \sum_i (f_i - f_.)\lambda_i \ .$$

Thus

$$E(\sum_i (\hat{\phi}_i - \hat{\phi}_.)^2 = \sum_i \lambda_i^2 + E(\sum_i (f_i - f_.)^2)$$

$$= \sum_i \lambda_i^2 + E(\sum_i f_i^2 - b\, f_.^2)$$

$$= \sum_i \lambda_i^2 + \sum_i \mathrm{Var}(f_i) - \frac{1}{b} \mathrm{Var}(\sum_i f_i)$$

$$= \sum_i \lambda_i^2 + \sum_i \mathrm{Var}(f_i)(1 - b^{-1})$$

$$= \sum_i \lambda_i^2 + b^2(1 - b^{-1})\ \mathrm{Var}(\hat{\phi}) \qquad (5.24)$$

as from (5.20)

$$\hat{\phi} = \hat{\phi}_. \text{ and } \mathrm{Var}(\phi) = \frac{1}{b^2} \sum_i \mathrm{Var}(f_i) \ .$$

From (5.24) we get

$$E(\frac{s^2}{b}) = b^{-1}(b-1)^{-1} \sum_i \lambda_i^2 + \mathrm{Var}(\hat{\phi}) \ .$$

Q.E.D.

If all the block-treatment interactions γ_{ij} are zero, then s^2/b is an unbiased estimate of $\mathrm{Var}(\hat{\phi})$.

Let us now find the mathematical expectations of $Q_1/(b-1)$, $Q_2/(\nu-1)$ and $Q_3/(b-1)(\nu-1)$, defined in Table 5.2, under the randomization model. We shall assume here that the e_{ij} are uncorrelated.

Now, since $E(e_{ij}) = 0$ for all i, j and the δ_{ijk} are independent of the e_{ijk}, we get from Lemma 5.1

$$\mathrm{Cov}(e_{ij}, e_{i'j'}) = E(\sum_k \delta_{ijk} e_{ijk}, \sum_{k'} \delta_{i'j'k'} e_{i'j'k'})$$

$$= \sum_k \sum_{k'} E(\delta_{ijk}\delta_{i'j'k'}) E(e_{ijk} e_{i'j'k'})$$

$$= \sum_k \sum_{k'} \delta_{ii'}\delta_{jj'}\delta_{kk'} \mathrm{Var}(e_{ijk}) E(\delta_{ijk}\delta_{i'j'k'})$$

$$= \delta_{ii'}\delta_{jj'} \nu^{-1} \sum_k \mathrm{Var}(e_{ijk}) . \tag{5.25}$$

From (5.25) we conclude that if the e_{ijk} are uncorrelated, then the e_{ij} are uncorrelated.

Let $\sigma_B^2 = \frac{1}{b-1} \sum_i b_i^2$, $\sigma_T^2 = \frac{1}{\nu-1} \sum_j t_j^2$,

$$\sigma_{BT}^2 = \frac{1}{(b-1)(\nu-1)} \sum_i \sum_j \gamma_{ij}^2 ,$$

$$\sigma_e^2 = \frac{1}{\nu b} \sum_i \sum_j \mathrm{Var}(e_{ij})$$

$$= \frac{1}{\nu^2 b} \sum_i \sum_j \sum_k e_{ijk}^2 , \tag{5.26}$$

$$\sigma_U^2 = \frac{1}{b(\nu-1)} \sum_i \sum_k (\mu_{i.k}-\mu_{i..})^2$$

$$= \frac{1}{b(\nu-1)} \sum_i \sum_k \theta_{ik}^2 , \text{ (say)},$$

$$= \frac{1}{b(\nu-1)^2} \sum_i \sum_j \sum_k (\mu_{ijk}-\mu_{i.k}-\mu_{ij.} + (\mu_{i..})^2$$

$$\sigma_{TU}^2 = \frac{1}{b(\nu-1)^2} \sum_i \sum_j \sum_k \nu_{ijk}^2 , \text{ (say)}.$$

Then under the assumption that the technical errors are uncorrelated one can obtain after straightforward but tedious calculations the following lemma.

Lemma 5.3.

$$E\left(\frac{Q_1}{b-1}\right) = \nu\sigma_B^2 + \nu^{-1}(\nu-1)\,\sigma_{TU}^2 + \sigma_e^2 ,$$

$$E\left(\frac{Q_2}{\nu-1}\right) = b\,\sigma_T^2 + \sigma_U^2 + \nu^{-1}(\nu-2)\,\sigma_{TU}^2 + \sigma_e^2 ,$$

$$E\left(\frac{Q_3}{(b-1)(\nu-1)}\right) = \sigma_{BT}^2 + \sigma_U^2 + \nu^{-1}(\nu-2)\,\sigma_{TU}^2 + \sigma_e^2 .$$

From Lemma 5.3 if $\sigma_T^2 = \sigma_{BT}^2 = 0$. Then

$$E\left(\frac{Q_2}{\nu-1}\right) = E\left(\frac{Q_3}{(b-1)(\nu-1)}\right)$$

even if $\sigma_{TU}^2 \neq 0$.

(b) Testing. (Permutation test). Under the randomization model exact tests of hypotheses in the randomized block design exist whenever the joint distribution of the observations under the hypothesis remains invariant under the group of permutations of the observations.

We shall first give the general idea of the permutation test of a hypothesis H. Permutation tests are constructed when the distribution of the random observations is invariant under a group of permutations of these observations. Let $\underline{X} = (X_1,\ldots,X_N)'$ be a vector of random observations and let $\underline{x} = (x_1,\ldots,x_N)'$ be one of its values. Suppose there exists a group G of L permutations of the N components of $\underline{X}$ with the following invariant property.

Let $S(\underline{x}) = \{g\underline{x}\colon g \in G\}$ be the set of L sample observations generated by $\underline{x}$, applying the group of permutations G to $\underline{x}$. The required symmetry property is that, given the condition that $\underline{X} \in S(\underline{x})$, the conditional probability under the hypothesis H that $\underline{X}$ takes on any particular one of the L values in $S(\underline{x})$ is 1/L for each of the L elements in $S(\underline{x})$ and this holds for all values $\underline{x}$ of the random vector $\underline{X}$.

A permutation test of H at a level of significance α is defined, for each possible set $S(\underline{x})$, by a subset $S'(\underline{x})$ of samples whose number of elements is less than or equal to αL and the hypothesis H is rejected if and only if the observed sample $\underline{x}$ is in $S'(\underline{x})$. The subset $S'(\underline{x})$ is generally defined by means of a test statistic. In most situations the test statistics considered are analogous to those employed in the analysis of variance technique with the added assumption of normality and independence of $X_1,\ldots,X_N$. For a thorough treatment of the permutation test the reader is referred to Scheffe (1943, 1967), Hotelling (1952) or Lehmann (1959).

Example 5.2. Consider an one-way classification with 3 observations from each of 2 populations corresponding to two treatments. Denote by x_{ij}, the jth random observation from the ith population

$x_{11} = 9,\ x_{12} = 4,\ x_{13} = 8,\ x_{21} = 7,\ x_{22} = 6,\ x_{23} = 5$, and let $\underline{x}=(x_{11},\ldots,x_{13}, x_{21},\ldots,x_{23})'$. Suppose that we are interested by means of sample observations (9,4,8)' and (7,6,5)' in testing the hypothesis H of the equality of treatment effects. We shall now find the permutation test based on the statistic

$$|T(\underline{x})| = \left|\sum_{j=1}^{3} x_{1j} - \sum_{j=1}^{3} x_{2j}\right| .$$

There are twenty different ways of dividing 6 objects into groups of 3. Thus in $S(\underline{x})$ there are 20 different elements which give different values of $T(\underline{x})$.

For the observed $\underline{x}$, $T(\underline{x}) = 3$. Under H, from the Table 5.4 $P\{|T(\underline{x})| \geq 3\} = .70$.

If we assume that the X_{ij} are independently normally distributed and apply the analysis of variance technique to analyse the data we get Table 5.5. In both the cases, we accept the hypothesis H with high probability.

We shall now define permutation tests in the randomized block design. These tests do not depend on the assumption of normality, independence and the homogeneity of the error variance which are assumed in (5.2). Consider the randomized block design in which the observation x_{ij} on the jth treatment $(j=1,\ldots,v)$ in

Table 5.4. Coordinates of $\underline{x}$ for different permutations

x_{11}, x_{12}, x_{13}			x_{21}, x_{22}, x_{23}			$T(\underline{x})$
9	4	8	7	6	5	3
7	6	5	9	4	8	-3
9	4	7	8	6	5	1
8	6	5	9	4	7	-1
9	4	6	7	8	5	-1
7	8	5	9	4	6	1
9	4	5	7	6	8	-3
7	6	8	9	4	5	3
9	8	7	4	6	5	9
4	6	5	9	8	7	-9
9	8	6	4	7	5	7
4	7	5	9	8	6	-7
9	8	5	4	7	6	5
4	7	6	9	8	5	-5
9	7	6	4	8	5	5
4	8	5	9	7	6	-5
9	7	5	4	8	6	3
4	8	6	9	7	5	-3
9	6	5	4	8	7	1
4	8	7	9	6	5	-1

Table 5.5. Analysis of Variance

Sources of variation	d.f.	sum of squares	F
Treatments	1	1.5	.187
Within treatments	4	8.0	
Total	5	9.5	

the ith block $(i = 1,\ldots,b)$ has the following structure.

$$x_{ij} = \mu + \alpha_i + t_j + \gamma_{ij} + \eta_{ij} + e_{ij} \qquad (5.27)$$

where

$$\eta_{ij} = \sum_k \delta_{ijk}\eta_{ijk}, \quad e_{ij} = \sum_k \delta_{ijk}e_{ijk},$$

$$\sum_j t_j = 0, \quad \sum_i \alpha_i = 0, \quad \sum_j \gamma_{ij} = \sum_i \gamma_{ij} = 0 .$$

Under the hypothesis H_T: $t_1 = \dots = t_\nu = 0$, the distribution of x_{ij} does not possess this invariant property. However such an invariant property can be achieved if we refine H_T as the hypothesis H_T' that there is absolutely no difference between treatment effects in the sense that

(i) $\sigma_T^2 = \sigma_{TB}^2 = \sigma_{TU}^2 = 0$;

(ii) the joint distribution of the e_{ij} does not depend on how the treatments are assigned to the $b\nu$ experimental units. Under H_T'

$$x_{ij} = \mu + \alpha_i + \sum_k \delta_{ijk}(\theta_{ik} + u_{ik}),$$

where the θ_{ij} are defined in (5.26) and the u_{ij} play the role of e_{ijk} which are now assumed not to depend on how the treatments are assigned to the experimental units. By definition, $E(U_{ik})=0$, but, otherwise the joint distribution of the U_{ik} is arbitrary.

Let G be the set of all permutations within blocks and there are obviously $(\nu!)^b$ elements in G. Let $\underline{x} = (x_{11},\dots,x_{1\nu},\dots,x_{b1},\dots,x_{b\nu})'$ and let

$$S(\underline{x}) = \{g\underline{x}: g \in G\}$$

be the set of all $(\nu!)^b$ samples obtained from $\underline{x}$ by applying all permutations in G. Let us first show that under H_T' and the randomization incorporated in the experiment all samples in $S(\underline{x})$ have the same conditional probability given that the given sample is in $S(\underline{x})$.

Let $\underline{X}$ be the random vector that takes on the value $\underline{x}$ in the experiment. Let

$$X_{ij} = \sum_k \delta_{ijk}Z_{ik} \tag{5.28}$$

where the Z_{ik} are random variables taking on the values z_{ik}, given by,

$$z_{ik} = \mu + \alpha_i + \theta_{ik} + u_{ik},$$

that is, z_{ik} is the observation on the (i,k) experimental unit and the δ_{ijk} are defined in (5.17). Denote by

$$\underline{Z} = (Z_{11},\ldots,Z_{1\nu},\ldots,Z_{b1},\ldots,Z_{b\nu})'$$

and by $\underline{z}$ the value that $\underline{Z}$ takes on in the experiment. A particular assignment of treatments to the experimental units in the randomization determines δ_{ijk} and hence $\underline{X}$ in terms of $\underline{Z}$ by the linear transformation (5.28), and such assignments have equal probability. In other words all coordinates of $\underline{X}$ are formed from the coordinates of $\underline{Z}$ by one of the permutations in G and all permutations have equal probability. Since for $\underline{Z} \varepsilon S(\underline{z})$, $S(\underline{Z}) = S(\underline{z})$, we can say that given $\underline{z} \varepsilon S(\underline{z})$ all $\underline{X}$ in $S(\underline{z})$ have equal probability. Finally $S(\underline{x}) = S(\underline{z})$ and $\underline{Z} \varepsilon S(\underline{z})$ if and only if $\underline{X} \varepsilon S(\underline{x})$. Hence we conclude that given $\underline{X} \varepsilon S(\underline{x})$ all $\underline{X}$ in $S(\underline{x})$ have the same conditional probability.

We now need to find out the subset $S'(\underline{x})$ such that H_T' will be rejected if the observed $\underline{x}$ falls in $S'(\underline{x})$ through a test criterion. Let us consider the test statistic $F = (b-1)\,Q_2/Q_3$ which was used in Table 5.2 for testing H_T. Since

$$q_2+q_3 = \sum_i \sum_j x_{ij}^2 - b\nu\, x_{..}^2 - \sum_i \nu(x_{i.}-x_{..})^2,$$

q_2+q_3 has a constant value for all elements in $S(\underline{x})$. It follows that F is a strictly increasing function of

$$Q_2 = \sum_j b(X_{.j}-X_{..})^2$$

$$= \frac{1}{b} \sum_j (\sum_i X_{ij})^2 - b\nu X_{..}^2 ,$$

and hence of $\sum_j (\sum_i X_{ij})^2$. So the permutation test based on the test statistic

$$T(\underline{X}) = \sum_j (\sum_i X_{ij})^2$$

is equivalent to the permutation test based on the statistic F. Since $T(\underline{X})$ is invariant by any permutation of treatments, the $(\nu!)^b$ values of $T(\underline{X})$ are equal in sets of $\nu!$. Thus there are $(\nu!)^{b-1}$ values assumed by $T(\underline{X})$ with equal probability under H_T'. In practice $(\nu!)^{b-1}$ is quite large and the calculation of these values to find the distribution of $T(\underline{X})$ under H_T' is often impracticable and we need to find an approximate distribution of $T(\underline{X})$ under H_T'.

The test based on F is also equivalent to the test based on the $V = Q_2/(Q_1+Q_2)$ for testing H_T and V has central beta distribution under H_T (see Giri, 1975). We shall consider here the permutation test based on the statistic V (which is equivalent to basing on F) instead of F itself. The main reason for this is that V unlike F has a constant denominator under the permutations in G. We shall find the approximate permutation distribution of V under H_T' by equating the first two moments of V with the first two moments of a central beta distribution.

Case (i). In addition to H_T' let us assume here that there are no technical errors, that is, $\sigma_e^2 = 0$, hence all $u_{ik} = 0$. From Lemma 5.3, under H_T' and the additional assumption $\sigma_e^2=0$.

$$E(Q_2) = (\nu-1)\,\sigma_U^2, \tag{5.29}$$

$$E(Q_3) = (b-1)(\nu-1)\,\sigma_U^2 .$$

The constant value of $Q_2 + Q_3$ in the set $S(\underline{x})$ may be calculated as

$$\begin{aligned} q_2 + q_3 &= E(Q_2+Q_3) \\ &= b(\nu-1)\,\sigma_U^2 \\ &= \sum_i \sum_k \theta_{ik}^2 . \end{aligned} \tag{5.30}$$

From (5.29) under H_T' and $\sigma_e^2 = 0$,

$$E(V) = (q_2+q_3)^{-1} E(Q_2) = b^{-1} \tag{5.31}$$

The Var(V) has been calculated by Welch (1937) and verified by Pittman (1937) as

$$\mathrm{Var}(V) = 2b^{-2}(\nu-1)^{-1}\left(1-b^{-2}\,\sigma_U^{-4} \sum_i \sigma_{U.i}^4\right) \tag{5.32}$$

where $\sigma_{U.i}^2 = (\nu-1)^{-1} \sum_k \theta_{ik}^2$. Since it involves complicated computations we refrain from presenting it here.

Under the present assumption the block variances can be calculated without error from the observations x_{ij}. Since with $u_{ik} = 0$, $\theta_{ik} = z_{ik} - z_{i.}$ and hence

$$\sigma_{U.i}^2 = (\nu-1) \sum_k (z_{ik}-z_{i.})^2$$

$$= (\nu-1) \sum_j (x_{ij} - x_{i.})^2 .$$

From (5.22) - (5.23) (5.33)

$$\text{Var}(V) = \frac{2(b-1)}{b^3(\nu-1)} \left(1 - \frac{\rho^2}{b}\right) , \qquad (5.34)$$

where

$$\frac{\rho^2}{b} = \sum_i (\sigma^2_{U.i} - \sigma^2_U)^2 / b(b-1)\sigma^4_U .$$

$$= \frac{1}{b-1} \left[\frac{b \sum_i s_i^2}{(\sum_i s_i)^2} - 1 \right] , \qquad (5.35)$$

$$s_i^2 = \sum_j (x_{ij} - x_{i.})^2 .$$

Since under the normality assumption V is distributed as central beta we shall assume that under H'_T and $\sigma^2_e = 0$ the distribution of V under the randomization model is also central beta with parameters (or degrees of freedom) ν_1, ν_2, say. The values of ν_1, ν_2 are obtained by equating the mean and the variance of V with the mean and variance of the assumed beta distribution. Since the mean and the variance of a central beta random variable X with parameters ν_1, ν_2 are

$$E(X) = \frac{\nu_1}{\nu_1+\nu_2} , \quad \text{Var}(X) = \frac{2\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+2)} , \qquad (5.36)$$

from (5.31), (5.34) and (5.36) we get

$$\nu_1 = \phi(\nu-1) , \quad \nu_2 = \phi(\nu-1)(b-1) , \qquad (5.37)$$

$$\phi = \left(1 - \frac{\rho^2}{b}\right)^{-1} - 2b^{-1}(\nu-1)^{-1} .$$

The approximate permutation test of H'_T on the basis of the statistic V rejects H'_T whenever the observed value of V for the observed sample $\underline{x}$ is greater than or equal to the $(1-\alpha)$ percent point of the central beta distribution with parameters ν_1, ν_2 given in (5.37).

This approximation of beta distribution on the basis of the first two moments of V also fits quite closely with respect to th

3rd and the 4th moment of V unless $b^{-1}\rho^2$ nears 1. Ordinarily the factor ϕ deviates significantly from unity and the deviation is positive. Since for $1-\alpha \geq .90$, $\nu_2 > 2$, $F_{1-\alpha,\nu_1,\nu_2}$ is a decreasing function of ν_1 and ν_2, the values of the F-statistic which miss significance with the unadjusted degrees of freedom may achieve it with adjusted degrees of freedom given in (5.37). Thus the permutation test is more sensitive than the unadjusted normal theory test. Note that if a random variable X has central beta distribution with ν_1, ν_2 degrees of freedom then $\nu_2 X/\nu_1(1-X)$ has central F-distribution with ν_1, ν_2 degrees of freedom and is a strictly increasing function of X.

Case (ii). The U_{ik} have an arbitrary distribution with $E(U_{ik}) = 0$.

With respect to the conditional distribution of the X_{ij} given the values u_{ik} of U_{ik}, from (5.37), we can write under H'_T

$$x_{ij} = \mu' + \alpha'_i + \sum_k \delta_{ijk}\theta'_{ik} , \tag{5.38}$$

where

$$\mu' = \mu + u_{..} , \qquad \alpha'_i = \alpha_i + u_{i.} - u_{..} ,$$

$$\theta'_{ik} = \theta_{ik} + u_{ik} - u_{i.} ,$$

$$\sum_i \alpha'_i = 0, \quad \sum_k \theta'_{ik} = 0 \quad \text{for all } i.$$

Hence the observations x_{ij} are distributed conditionally like those in (5.27) with $(\mu, \alpha_i, \theta_{ik})$ replaced by $(\mu', \alpha'_i, \theta'_{ik})$. Thus the conditional distribution of V under the randomization model (5.38) is the same as the distribution of V (in case i) where $u_{ik} = 0$ and θ_{ik} are replaced by θ'_{ik}. In particular in this conditional case the mean and the variance of V under H'_T are given by (5.31) and (5.34) respectively. Thus the approximation to this conditional permutational test is made by carrying out the normal theory beta test (or equivalently the F-test) with degrees of freedom adjusted as in (5.37) by the factor ϕ. Actually in this case ϕ is a random variable but has a constant value in the conditional case. However if the test conditionally controls the probability of type I error so that its size is approximately α, then the same will be true unconditionally. For further discussions and results in this case the reader is referred to Scheffe (1967) and Hotelling (1952).

5.2.5. Efficiency and the Optimality of Randomized Block Designs:

We shall now consider the efficiency and the optimality of the randomized block design relative to the completely randomized design in which the treatments are assigned to the bv experimental units subject to a restriction that each treatment occurs b times in such a way that all assignments have equal probability. We shall also assume here that all interactions between experimental units and treatments in the completely randomized design are zero. This will imply that in any grouping of experimental units into blocks the treatment-block interactions γ_{ij} and the treatment-unit interactions within a block ν_{ijk} in (5.26) are all zero, that is,

$$\sigma^2_{BT} = 0, \quad \sigma^2_{TU} = 0 , \tag{5.39}$$

furthermore, for all i,j,k, the technical errors e_{ijk} are additive, that is,

$$e_{ijk} = t_{ijk} + u_{ijk} , \tag{5.40}$$

where the t_{ijk} are associated with treatments and the u_{ijk} are associated with experimental units satisfying

$$E(t_{ijk}) = 0, \quad E(u_{ijk}) = 0 \tag{5.41}$$

and the t_{ijk} and the u_{ijk} are independent of each other and of the δ_{ijk}. This permits us to write

$$\mathrm{Var}(e_{ijk}) = \sigma^2_{t.j} + \sigma^2_{u.ik} , \tag{5.42}$$

where

$$\mathrm{Var}(t_{ijk}) = \sigma^2_{t.j} , \quad \mathrm{Var}(u_{ijk}) = \sigma^2_{u.ik} . \tag{5.43}$$

Note that for convenience we are using here the same notation for the random variable and its value.

The Equation (5.15) for the completely randomized design becomes

$$x_{ij} = \mu + \alpha_i + t_j + \eta_{ij} + e_{ij} \tag{5.44}$$

where

$$\sum_i \alpha_i = 0 , \quad \sum_j t_j = 0 ,$$

$$\eta_{ij} = \sum_k \delta_{ijk}\eta_{ijk} = \sum_k \delta_{ijk}(\mu_{i.k}-\mu_{i..})$$

$$= \sum_k \delta_{ijk}\,\theta_{ik} ,$$

as all ν_{ijk} are zero, and

$$e_{ij} = \sum_k \delta_{ijk}e_{ijk} = \sum_k \delta_{ijk}(t_{ijk}+u_{ijk}) \tag{5.45}$$

Lemma 5.4.

(i) $\mathrm{Cov}(e_{ij},e_{i'j'}) = \delta_{ii'}\delta_{jj'}(\sigma^2_{t.i}+\sigma^2_{u.i})$, (5.46)

where $\sigma^2_{u.i} = \nu^{-1} \sum_k \sigma^2_{u.ik}$,

(ii) $\mathrm{Cov}(\eta_{ij},e_{i'j'}) = 0$, (5.47)

$$\text{(iii)}\ \mathrm{Cov}(\eta_{ij},\eta_{i'j'}) = \begin{cases} \delta_{ii'}(1-v^{-1})\ \sigma^2_{U.i} , & \text{if}\quad j = j' , \\ -\delta_{ii'}\ v^{-1}\ \sigma^2_{U.i} & \text{if}\quad j \neq j' , \end{cases} \tag{5.48}$$

where $\sigma^2_{U.i} = (\nu-1)^{-1} \sum_k \theta^2_{ik}$.

Proof. (i) Since $E(e_{ij}) = 0$,

$$\mathrm{Cov}(e_{ij},e_{i'j'}) = E(e_{ij},e_{i'j'})$$

$$= \sum_k \sum_{k'} E(\delta_{ijk}\delta_{i'j'k'})\ E(e_{ijk}e_{i'j'k'})$$

$$= \delta_{ii'}\delta_{jj'}\ v^{-1} \sum_k \mathrm{Var}(e_{ijk}) = \delta_{ii'}\delta_{jj'}\ v^{-1} \sum_k(\sigma^2_{t.i}+\sigma^2_{u.ik})$$

$$= \delta_{ii'}\delta_{jj'}\ (\sigma^2_{t.i} + \sigma^2_{u.i}) .$$

(ii) Since $E(e_{ijk}) = 0$,

$$\mathrm{Cov}(\eta_{ij},e_{i'j'}) = \sum_k \sum_{k'} E(\delta_{ijk}\delta_{i'j'k'})\ \theta_{ik}\ E(e_{i'j'k'})$$

$$= 0 .$$

(iii) From (5.45)

$$\text{cov}(\eta_{ij},\eta_{i'j'}) = E(\eta_{ij}\eta_{i'j'})$$

$$= \sum_k \sum_{k'} E(\delta_{ijk}\delta_{i'j'k'})\ \theta_{ik}\ \theta_{i'k'}$$

$$= \delta_{ii'} \sum_k \sum_{k'} E(\delta_{ijk}\delta_{i'j'k'})\ \theta_{ik}\theta_{ik'}\ .$$

Thus for $j = j'$

(5.45) $$\text{cov}(\eta_{ij},\eta_{i'j'}) = \delta_{ii'} \sum_k \sum_{k'} \nu^{-1}\ \delta_{kk'}\theta_{ik}\theta_{ik'}$$

$$= \delta_{ii'}\ \nu^{-1} \sum_k \theta^2_{ik} = \delta_{ii'}(1-\nu^{-1})\,\sigma^2_{U.i},$$

(5.46) and for $j \neq j'$.

$$\text{cov}(\eta_{ij},\eta_{i'j'})$$

(5.47) $$= \delta_{ii'}\ \nu^{-1}(\nu-1)^{-1} \sum_k \sum_{k'} (1-\delta_{kk'})\,\theta_{ik}\theta_{ik'}$$

$$= -\delta_{ii'}\ \nu^{-1}\ \sigma^2_{U.i}\ ,$$

(5.48) since $\sum_k \theta_{ik} = 0$. Q.E.D.

From (5.44)

$$x_{.j} = \mu + t_j + \eta_{.j} + e_{.j}\ .$$

Therefore, by Lemma 5.4, the variance of the estimate $\hat{\phi} = \sum_j c_j x_{.j}$ of the contrast $\phi = \sum_j c_j t_j$ with $\sum_j c_j = 0$ is

$$\text{Var}(\hat{\phi}) = \sum_j \sum_{j'} c_j c_{j'}\ [\text{Cov}(\eta_{.j},\eta_{.j'})+\text{Cov}(e_{.j},e_{.j'})]$$

$$= b^{-2} \sum_j \sum_{j'} \sum_i \sum_{i'} c_j c_{j'} [\text{Cov}(\eta_{ij},\eta_{i'j'})+\text{Cov}(e_{ij},e_{i'j'})]$$

$$= b^{-1}(\sigma_U^2+\sigma_u^2) \sum_j c_j^2 + \sum_j c_j^2\ \sigma^2_{t.j}\ , \qquad (5.49)$$

where

$$\sigma_U^2 = b^{-1} \sum_i \sigma^2_{U.i}\ ,$$

$$\sigma_u^2 = b^{-1} \sum_i \sigma^2_{u.i} = b^{-1}\nu^{-1} \sum_i \sum_k \sigma^2_{u.ik'} \qquad (5.50)$$

and σ_U^2 refers to the unit error within block, σ_u^2 refers to the components of the technical errors associated with units and $\sigma_{t.j}^2$ refers to the components associated with the jth treatment.

We now consider the completely randomized designs. For the sake of comparison of the completely randomized design with the randomized block design we shall, as in the case of the randomized block design, number the νb experimental units with two indices i,j as though they are grouped in b blocks of ν units each but of course no attention is paid for this fictitious blocking in the random assignment of treatments to experimental units. We shall also assume that the same experimental material is used for either design. For field experimentation the shape and the size of experimental units (plots) are exactly identical for both designs. Under the additivity assumption the observed measurement x_{ijk} with the treatment j on the plot k in the ith block can be written as

$$x_{ijk} = \mu + t_j + \beta_{ik} + t_{ijk} + u_{ijk} \qquad (5.51)$$

$$i = 1,\ldots,b,\ j = 1,\ldots,\nu$$

$$k = 1,\ldots,\nu\ ,$$

where $\beta_{ik} = \theta_{ik} + \alpha_i$ is a unit effect (not within the ith block like θ_{ik}) satisfying $\sum_i \sum_k \beta_{ik} = 0$.

Here

$$\tilde{\sigma}_U^2 = (b\nu-1)^{-1} \sum_i \sum_k \beta_{ik}^2 \qquad (5.52)$$

is a measure of the magnitude of the unit effect, to be distinguished from σ_U^2 in (5.26) which measures the unit effects within blocks. The $\tilde{\sigma}_U^2$ does not depend on how the experimental units are grouped but the σ_U^2 does. Now from (5.52)

$$(b\nu-1)\,\tilde{\sigma}_U^2 = \sum_i \sum_k \beta_{ik}^2 = \sum_i \sum_k (\alpha_i + \theta_{ik})^2 \qquad (5.52a)$$

$$= \nu(b-1)\,\sigma_\beta^2 + b(\nu-1)\,\sigma_U^2\ ,$$

and the observed jth treatment mean

$$x_{.j} = \mu + t_j + b^{-1} \sum_i \sum_k d_{ijk}(\beta_{ik} + t_{ijk} + u_{ijk}), \tag{5.53}$$

where the random variables d_{ijk} play the same role as the δ_{ijk} in the randomized block design and are defined by

$$d_{ijk} = \begin{cases} 1, & \text{if the jth treatment is applied to (i,k) unit,} \\ 0, & \text{otherwise,} \end{cases} \tag{5.54}$$

and from Lemma 5.1

$$E(d_{ijk}, d_{i'j'k'}) = \begin{cases} \dfrac{b-\delta_{jj'}}{(\nu b-1)\nu}, & \text{if } (i,k) \neq (i',k'), \\ 0, & \text{if } (i,k) = (i',k'). \end{cases} \tag{5.55}$$

From (5.49) the variance of the unbiased estimator $\tilde{\phi} = \sum_j c_j x_{.j}$ of the contrast $\sum_j c_j t_j$ ($\sum_j c_j = 0$) is

$$\mathrm{Var}(\tilde{\phi}) = b^{-1} [(\tilde{\sigma}_U^2 + \sigma_u^2) \sum_j c_j^2 + \sum_j c_j^2 \sigma_{t.j}^2)] \tag{5.56}$$

From (5.49) and (5.56) we conclude that the randomized block design is more efficient if

$$\sigma_U^2 < \tilde{\sigma}_U^2 \tag{5.57}$$

or equivalently $\sigma_B^2 > \nu^{-1} \sigma_v^2$ and is independent of the contrast studied. The efficiency of the randomized block design relative to the completely randomized design is defined by (from (5.49) and (5.56))

$$\varepsilon = \frac{(\tilde{\sigma}_U^2 + \sigma_u^2) \sum_j c_j^2 + \sum_j c_j^2 \sigma_{t.j}^2}{(\sigma_U^2 + \sigma_u^2) \Sigma c_j^2 + \sum_j c_j^2 \sigma_{t.j}^2} \tag{5.58}$$

which is dependent on the individual contrast studied unless the $\sigma_{t,j}^2$ are all equal to σ_t^2 (say). In such cases, using (5.52),

$$\varepsilon = \frac{\tilde{\sigma}_U^2 + \sigma_u^2 + \sigma_t^2}{\sigma_U^2 + \sigma_u^2 + \sigma_t^2} = \frac{\tilde{\sigma}_U^2 + \sigma_e^2}{\sigma_U^2 + \sigma_e^2} \tag{5.59}$$

$$= \frac{(b-1)(\nu\sigma_B^2 + \sigma_e^2) + b(\nu-1)(\sigma_U^2 + \sigma_e^2)}{(b\nu-1)(\sigma_U^2 + \sigma_e^2)}$$

where $\sigma^2 = \sigma_t^2 + \sigma_e^2$. From Lemma 5.3 and equation (5.39) and the fact that $\sigma_{t.j}^2 = \sigma_t^2$ for all j, we get

$$E\left(\frac{Q_1}{b-1}\right) = \nu\sigma_B^2 + \sigma_e^2 ,$$

$$E\left(\frac{Q_3}{(b-1)(\nu-1)}\right) = \sigma_U^2 + \sigma_e^2 . \tag{5.60}$$

From (5.59)-(5.60) the estimate is given by

$$\hat{\varepsilon} = \frac{q_1 + \frac{b}{(b-1)} q_3}{\frac{(b\nu-1)}{(b-1)(\nu-1)} q_3} . \tag{5.61}$$

Cochran (1938, 1940) concluded from the analysis of experiments carried out at the Rothamsted experimental station and at its associated centres between 1927-1934 that the mean error sum of squares (the estimate of σ^2) for the randomized block design was sixty percent of the mean error sum of squares for the completely randomized design. Thus assuming that all experiments have equal cost factors it would require 10 replicates of a completely randomized design to obtain an amount of information equal to that from a randomized block design with 6 replicates.

5.3. LATIN SQUARES DESIGN

5.3.0. Introduction:

In the randomized block design the treatments are assigned to blocks at random such that each treatment must occur in each block rather than assigning them at random over the entire set of experimental units as in a completely randomized design. For a latin square design the experimental area is divided into rows and columns each having the same number of experimental units (equal to the number of treatments) and the treatments are allocated to rows and columns in such a way that each treatment occurs once and only once in each row and in each column. Thus in the latin square design the treatments are grouped into replicates in two ways once in rows and once in columns. Whereas in the randomized block design, only the block variation is eliminated from the within treatment variation, in the latin square design, row and column variations are eliminated from the within treatment variation. As a result in the latin square design the estimate of the error variance can be considerably reduced.

A latin square design is an incomplete 3-way layout in which each of the 3 factors - rows, columns and treatments - is at ν levels and observations on only ν^2 of the ν^3 possible treatment combinations (each treatment combination contains one level of each factor) are taken.

Latin square designs were originated in agricultural experimentation by Fisher (1926). Now-a-days they are used in industrial, laboratory, greenhouse, field, medical, marketing and sociological experimentations, from comparing a group of fertilizers to testing biological assays, from testing different brands of tea to comparing patients in the hospital, etc. Despite its popularity it is only good if the number of treatments is not large.

5.3.1. Latin Squares: A latin square of order p is an arrangement of p symbols in p^2 cells arranged in p rows and p columns such that each symbol occurs once and only once in each row and in each column.

Example 5.3.

$$\begin{array}{cccc} A & B & C & D \\ B & C & D & A \\ C & D & A & B \\ D & A & B & C \end{array}$$

is a latin square of order 4 with the latin letters A, B, C, D as symbols.

A latin square is in the standard form if the symbols in the first row and in the first column are in the natural order.

Two standard latin squares are called conjugate if the rows of one are the columns of the other.

A latin square is called *self-conjugate* if its arrangement in rows and in columns are the same. The latin square in Example 5.3 is self-conjugate.

A set of all latin squares obtained from a single latin square by permuting its rows, columns and symbols is called a *transformation set*.

From a standard latin square of order p we may obtain p!(p-1)! different squares by making the p! permutations of columns and (p-1)! permutation of rows which leave the first row in place. Thus the number of different latin squares of order p in a transformation set is p!(p-1)! times the number of standard latin squares in the set. For example, for a latin square of order 2, there is only one standard square. The same is true for a latin square of order 3. For a latin square of order 4 there are 4 standard squares and for a latin square of order 5 there are 25 standard squares. This number increases rapidly as the order increases. Fisher and Yates (1934, 1948) obtained the standard squares of order 4 and 5 and the transformation set of order 6. Norton (1939) has tabulated the 562 sets from which it is possible to obtain the 16,927,968 standard latin squares of order 7.

If two latin squares of the same order but with different symbols are such that when they are superimposed on each other every ordered pair of symbols (different) occurs exactly once, then they are called *orthogonal*.

A pair of orthogonal latin squares, one with latin symbols and the other with greek symbols forms a *greco-latin* square.

Example 5.4.

$$\begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{array} \qquad \begin{array}{cccc} \alpha & \gamma & \beta & \delta \\ \beta & \delta & \gamma & \alpha \\ \gamma & \alpha & \beta & \delta \\ \delta & \beta & \alpha & \gamma \end{array}$$

are two orthogonal latin squares of order 4 and they form a greco-latin square.

A set of latin squares of the same order is called a set of mutually orthogonal latin squares (also a hyper greco-latin square) if every pair in the set is orthogonal.

The total number of mutually orthogonal latin squares of order p is at most p-1. A complete set of orthogonal latin squares of any order except 6 exists. For further results in this context and the construction of orthogonal latin squares we refer to MacNeish (1922), Bose (1938), Mann (1942), Parker (1959), Bose, Shrikhande and Parker (1960), Bose and Shrikhande (1959a,b, 1960), Raghav Rao (1971), Das and Giri (1976).

5.3.2. Analysis of Latin Square Designs (One Observation Per Cell): In designing a scientific experiment in a latin square of the required order p, one latin square is selected at random from the set of all possible squares of order p. This is achieved by selecting a standard square from the set of all standard squares with equal probability and then randomizing all the rows, columns and symbols. For detailed instructions on randomization we refer to Fisher and Yates (1943).

In developing the analysis of variance technique under the assumption of normality, the probability model we consider is a conditional one, given the condition that a particular latin square has been selected for the experiment.

Suppose that we have selected a latin square design of order ν for comparing ν treatments. Let

$$x_{ijk}, \quad i = 1,\ldots,\nu \quad , \; j = 1,\ldots,\nu \; , \quad k = 1,\ldots,\nu \; ,$$

denote the observation on the kth treatment in the ith row and the jth block. The triplets (i,j,k) take on only the ν^2 values indicated by the particular latin square selected for the experiment. Assume that the corresponding random variables X_{ij} are independently normally distributed with

$$\mu_{ijk} = E(X_{ijk}) = \mu + \alpha_i + \beta_j + t_k \, , \tag{5.62}$$

$$\mathrm{Var}(X_{ijk}) = {}^2 \, ,$$

where α_i is the ith row effect satisfying $\sum_i \alpha_i = 0$, β_j is the jth column effect satisfying $\sum_j \beta_j = 0$ and t_k is the kth treatment effect satisfying $\sum_k t_k = 0$. In the notations of Section 3.8, $\{\alpha_i\}$ are called the main effects of rows, $\{\beta_j\}$ are called the main effects of columns and $\{t_k\}$ are called the main effects of treatments.

We can also write

$$x_{ijk} = E(X_{ijk}) + e_{ijk} \tag{5.63}$$

where the ν^2 random variables e_{ijk} are independently normally distributed with mean 0 and variance σ^2. We shall consider the following hypothesis

$$H_R\text{: all } \alpha_i = 0, \quad H_C\text{: all } \beta_j = 0, \quad H_T\text{: all } t_k = 0. \tag{5.64}$$

In the notations of Section 3.1 each hypothesis is a linear hypothesis with $r = \nu-1$, $s = 1+3\nu-3 = 3\nu-2$ and $N-s = \nu^2-3\nu+2$. The least square estimates of the parameters $(\mu, \alpha_i, \beta_j, t_k)$ under can be obtained from the following identity (with S denoting the ν^2 values of (i,j,k))

$$\sum_{(i,j,k)\varepsilon S} (x_{ijk} - \mu - \alpha_i - \beta_j - t_k)^2$$

$$= \sum_{(i,j,k)\varepsilon S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})^2 + \nu\sum_i (x_{i..}-x_{...}-\alpha_i)^2$$

$$+ \nu\sum_j (x_{.j.}-x_{...}-\beta_j)^2 + \nu\sum_k (x_{..k}-x_{...}-t_k)^2 + \nu^2(x_{...}-\mu)^2 \tag{5.65}$$

From (5.68) it follows that

$$\hat{\alpha}_i = x_{i..}-x_{...}, \quad \hat{\beta}_j = x_{.j.}-x_{...}, \quad \hat{t}_k = x_{..k}-x_{...},$$

$$\hat{\mu} = x_{...},$$

$$\hat{\mu}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{t}_k$$

and

$$\sum_{(i,j,k)\varepsilon S} (x_{ijk}-\hat{\mu}_{ijk})^2 = \sum_{(i,j,k)\varepsilon S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})^2$$

Under H_R we still have (5.66)

$$\hat{\hat{\beta}}_j = x_{.j.}-x_{...}, \quad \hat{\hat{t}}_k = x_{..k}-x_{...}, \quad \hat{\hat{\mu}} = x_{...},$$

$$\hat{\hat{\mu}}_{ijk} = \hat{\hat{\mu}} + \hat{\hat{\beta}}_j + \hat{\hat{t}}_k,$$

and hence

$$\hat{\mu}_{ijk}-\hat{\hat{\mu}}_{ijk}= x_{i..}-x_{...}\ . \tag{5.67}$$

Thus the best invariant test of level α of H_R rejects H_R whenever

$$(\nu-2)\,\nu \sum_i (x_{i..}-x_{...})^2 \Big/ \sum_{(i,j,k)\,\varepsilon S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})^2$$

$$\geq F_{1-\alpha,\,\nu-1,\,\nu^2-3\nu+2}. \tag{5.68}$$

Similarly the best invariant test of level α of H_C rejects H_C whenever

$$(\nu-2)\,\nu \sum_j (x_{.j.}-x_{...})^2 \Big/ \sum_{(i,j,k)\,\varepsilon S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})^2$$

$$\geq F_{1-\alpha,\,\nu-1,\,\nu^2-3\nu+2}\ , \tag{5.69}$$

and the best invariant test of level α of H_T rejects H_T whenever

$$(\nu-2)\,\nu \sum_k (x_{..k}-x_{..})^2 \Big/ \sum_{(i,j,k)\,\varepsilon S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{..})^2 \tag{5.70}$$

$$\geq F_{1-\alpha,\,\nu-1,\,\nu^2-3\nu+2}.$$

The analysis of variance table for the latin square design is given in Table 5.6.

Table 5.6. ANOVA for latin square design

Sources of variation	d.f.	Sum of squares	Expected mean sum of squares	F
Rows	$\nu-1$	$q_1=\sum_i \nu(x_{i..}-x_{..})^2$	$\sigma^2+\frac{\nu}{\nu-1}\sum_i \alpha_i^2$	$\nu(\nu-2)\frac{q_1}{q_4}$
Columns	$\nu-1$	$q_2=\sum_j \nu(x_{.j.}-x_{...})^2$	$\sigma^2+\frac{\nu}{\nu-1}\sum_j \nu_j^2$	$\nu(\nu-2)\frac{q_2}{q_e}$
Treatments	$\nu-1$	$q_3=\sum_k \nu(x_{..k}-x_{...})^2$	$\sigma^2+\frac{\nu}{\nu-1}\sum_k t_k^2$	$\nu(\nu-2)\frac{q_3}{q_4}$
Error	$\nu^2-3\nu+2$	$q_4=\sum_{(i,j,k)\,\varepsilon S}(x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})^2$	σ^2	
Total	ν^2-1	$\sum_{(i,j,k)\,\varepsilon S}(x_{ijk}-x_{...})^2$		

In case the F-value corresponding to H_T is significant at a given level of significance α one has to use the multiple comparison method discussed in Chapter 3 to divide the entire set of treatments into different subgroups such that the treatments in the same subgroup have the same treatment effects and those in different subgroups have different treatment effects.

Example 5.5. In order to study the effects of varying doses of superphosphate on the yield of sugarcane crop under unirrigated condition an experiment was conducted at the sugarcane research station, Rayagude, Orissa, India. The treatments are

T_1 = control (zero lbs of P_2O_5 as superphosphate),

T_2 = 40 lbs of P_2O_2 per acre as superphosphate,

T_3 = 80 lbs of P_2O_2 per acre as superphosphate,

T_4 = 120 lbs of P_2O_2 per acre as superphosphate,

T_5 = 160 lbs of P_2O_2 per acre as superphosphate.

Plot size = 39' × 39'

Design: Latin square design of order 5.

The layout and the yield in kilograms within brackets sugarcane per plot is given below

T_2 (269)	T_4 (262)	T_5 (322)	T_3 (320)	T_1 (202)
T_4 (164)	T_2 (248)	T_3 (200)	T_1 (258)	T_5 (229)
T_5 (210)	T_1 (159)	T_4 (185)	T_2 (224)	T_3 (225)
T_1 (217)	T_2 (257)	T_2 (210)	T_5 (158)	T_4 (151)
T_1 (189)	T_5 (230)	T_1 (220)	T_4 (218)	T_2 (205)

Table 5.7. ANOVA table for the sugarcane data

Sources of variations	d.f	sum of squares	F
Rows	4	19,544.64	
Column	4	4,145.84	
Treatments	4	5,989.84	1.28
Error	12	17,476.72	
Total	24	47,157.04	

Since $F_{1-\alpha,4,12} = 2.48$, if $\alpha = 0.10$, we accept H_T, the hypothesis of the equality of treatment effects with high probability.

5.3.3. Missing Observations in the Latin Square Design: Allan and Wishart (1930), and Yates (1933) presented a formula for estimating one missing observation x_{ijk} in a latin square design of order p, given by

$$\hat{x}_{ijk} = \frac{\nu(x_{i..}+x_{.j.}+x_{..k}) - 2k_{...}}{(\nu-1)(\nu-2)} \tag{5.71}$$

As in the case of the randomized block design this estimate has been obtained by minimizing the error sum of squares, in the analysis of variance table of the Latin square design, with respect to the missing observation. In the original layout we may replace the missing observation by its estimate and perform the usual analysis of variance with the single exception that for the datum estimated 1 degree of freedom is subtracted from the actual degrees of freedom $\nu^2-3\nu+2$ of the sum of squares for errors (q_4 in Table 5.3.). As pointed out in connection with the randomized block design, the F-test obtained by this approximate approach will have an upward bias. However, if the F-value is non-significant we stop further investigation, but if the F-value is significant we cannot be sure that it is due to treatments and not due to the bias.

To obtain the exact test of H_T: all $t_k = 0$ the following procedure is adopted. The experiment is first regarded as a two-way classification in rows and columns with one missing observation. The minimum error sum of squares ω (say) for this classification (with respect to the missing observation) is obtained. Its degrees of freedom is $(\nu-1)^2-1 = \nu^2-2\nu$. The quantity $(\omega-q_4)$ is the exact treatment sum of squares with $(\nu-1)$ degrees of freedom. The exact test of H_T with one missing observation is to reject H_T if

$$\frac{\nu^2-3+1}{\nu-1}\left(\frac{\omega-q_4}{q_4}\right) \geq F_{1-\alpha,\nu-1,\nu^2-3\nu+1}$$

Yates (1933) has given an iterative method for estimating several missing observations in the latin square design. Bartlett (1937) suggested a procedure of inserting a "one" for the missing value and zero otherwise and performing a covariance analysis with the zeros and the one as the concomitant variable.

If more than one observation is missing the same procedure is followed except that a covariance analysis with more than one concomitant variates is performed. Nair (1940) used the Bartlett procedure for analysing a latin square design with several missing observations. DeLury (1946) summarizes most of the available results for handling missing observations in a latin square or sets of latin squares.

5.3.4. *Analysis of Covariance*: The analysis of covariance for the latin square design is a straightforward extension of the results on the analysis of covariance of the two-way classification in Chapter 4 or of the randomized block design in Section 5.2.

The linear model for the observation x_{ijk} from a latin square design of order ν with a single concomitant variate z_{ijk} is given by

$$x_{ijk} = \mu + \alpha_i + \beta_j + t_k + \gamma z_{ijk} + e_{ijk}, \tag{5.72}$$

where γ is the regression coefficient.

Let

$$\begin{aligned} R_{xz} &= \sum_i \nu(x_{i..}-x_{...})(z_{i..}-z_{...}) \\ C_{xz} &= \sum_j \nu(x_{.j.}-x_{...})(z_{.j.}-z_{...}), \end{aligned} \tag{5.73}$$

$$T_{xz} = \sum_{k} \nu(x_{..k}-x_{...})(z_{..k}-z_{...}),$$

$$E_{xz} = \sum_{(i,j,k)\in S} (x_{ijk}-x_{i..}-x_{.j.}-x_{..k}+2x_{...})(z_{ijk}-z_{i..}-z_{.j.}-z_{..k}+2z_{...}),$$

and let R_{xx}, R_{zz} be the values of R_{xz} when $z = x$, $x = z$ respectively. With similar interpretations for C_{xx}, C_{zz}, etc. the analysis of covariance for testing hypothesis H_R, H_C and H_T is given in Table 5.8. If any one of the tabulated F-values are significant we can determine by the multiple comparison methods which of the contrasts in the parameters concerned are responsible for the rejection.

5.3.5. Randomization Model

(a) Estimation. We have already considered the randomization model for the randomized block design. We shall now consider the same for the latin square design. For other designs to be discussed in this chapter we shall merely state the results on permutation in relevant places. The randomization model that we shall adopt for the latin square design will be valid regardless of how we choose the transformation set containing the latin square actually used in the experiment provided that all squares in the transformation set had the same probability of being selected for the experiment. Let x_{ijk} denote the observation of the kth treatment in the ith row and the jth column and let X_{ijk} be the corresponding random variable with

$$E(X_{ijk}) = \mu_{ijk} . \tag{5.74}$$

Then we can write

$$x_{ijk} = \mu_{ijk} + e_{ijk} \tag{5.75}$$

where the e_{ijk} are technical errors with $E(e_{ijk}) = 0$. To simplify the presentation we shall make more restrictive assumptions about the e_{ijk} than what we have done for the randomized block design. We shall assume here that the e_{ijk} are independently distributed with mean 0 and the same variance σ_e^2. Using the terminology of the 3-way classification in Section(3.8) with factors A(rows),

Table 5.8 Analysis of covariance for the latin square design.

Source of variation	d.f.	sum of products			Adjusted sum of squares		F
		xx	xz	zz	d.f.	sum of squares	
Rows	$\nu-1$	R_{xx}	R_{xz}	R_{zz}	$\nu-1$	$q_1 = q'_1 - q_4$	$\frac{\nu^2-3\nu+1}{\nu-1}\,\frac{q_1}{q_4}$
Columns	$\nu-1$	C_{xx}	C_{xz}	C_{zz}	$\nu-1$	$q_2 = q'_2 - q_4$	$\frac{\nu^2-3\nu+1}{\nu-1}\,\frac{q_2}{q_4}$
Treatments	$\nu-1$	T_{xx}	T_{xz}	T_{zz}	$\nu-1$	$q_3 = q'_3 - q_4$	$\frac{\nu^2-3\nu+1}{\nu-1}\,\frac{q_3}{q_4}$
Error	$\nu^2-3\nu+2$	E_{xx}	E_{xz}	E_{zz}	$\nu^2-3\nu+1$	q_4	
Rows & Error	$\nu^2-2\nu+1$					$q'_1 = (R_{xx}+E_{xx}) - \frac{(R_{xz}+E_{xz})^2}{R_{zz}+E_{zz}}$	
Columns & Error	$\nu^2-2\nu+1$					$q'_2 = (C_{xx}+E_{xx}) - \frac{(C_{xz}+E_{xz})^2}{C_{zz}+E_{zz}}$	
Treatments & Error	$\nu^2-2\nu+1$					$q'_3 = (T_{xx}+E_{xx}) - \frac{(T_{xz}+E_{xz})^2}{(T_{zz}+E_{zz})}$	

B(columns) and C(treatments; we can write in terms of the grand mean, main effects of A, B, C, two factor interactions of A, B, C and the three factor interaction of A, B, C

$$x_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^C + \alpha_{ij}^{AB} + \alpha_{ik}^{AC} + \alpha_{ij}^{BC} + \alpha_{ijk}^{ABC} + e_{ijk}, \tag{5.76}$$

where $i = 1, \ldots, \nu$, $j = 1, \ldots, \nu$, $k = 1, \ldots, \nu$ and

$$\begin{aligned} &\sum_i \alpha_i^A = \sum_j \alpha_j^B = \sum_k \alpha_k^C = 0, \\ &\sum_i \alpha_{ij}^{AB} = \sum_j \alpha_{ij}^{AB} = \sum_i \alpha_{ik}^{AC} = \sum_k \alpha_{ik}^{AC} = \sum_j \alpha_{jk}^{BC} = \sum_k \alpha_{jk}^{BC} = 0, \\ &\sum_i \alpha_{ijk}^{ABC} = \sum_j \alpha_{ijk}^{ABC} = \sum_k \alpha_{ijk}^{ABC} = 0, \end{aligned} \tag{5.77}$$

It may be noted that in Section 5.3.2 we have taken $\alpha_i^A = \alpha_i$, $\alpha_j^B = \beta_j$, $\alpha_k^C = t_k$ and all interactions to be zeros.

Let the randomization as in the case of the randomized block design generate random variable δ_{ijk} given by

$$\delta_{ijk} = \begin{cases} 1, & \text{if the kth treatment occurs in the ith row} \\ & \text{and in the jth column ,} \\ 0, & \text{otherwise ,} \end{cases} \tag{5.78}$$

such that the δ_{ijk} are independent of the e_{ijk}.

Then, from (5.76)

$$\begin{aligned} x_{ij.} &= \sum_k \delta_{ijk} x_{ijk}, \quad x_{i.k} = \sum_j \delta_{ijk} x_{ijk}, \\ x_{.jk} &= \sum_i \delta_{ijk} x_{ijk}, \end{aligned} \tag{5.79}$$

where the x_{ijk} are to be replaced by (5.76) and the $x_{ij.}$ is the observation in the ith row and in the jth column, etc.

The observed means for the ith row, the jth column and the kth treatment are, respectively,

$$\begin{aligned} x_{i..} &= \frac{1}{\nu} \sum_j \sum_k \delta_{ijk} x_{ijk}, \quad x_{.j.} = \frac{1}{\nu} \sum_i \sum_k \delta_{ijk} x_{ijk}, \\ x_{..k} &= \frac{1}{\nu} \sum_i \sum_j \delta_{ijk} x_{ijk}, \end{aligned} \tag{5.80}$$

and the observed general mean is

$$x_{\cdots} = \frac{1}{\nu^2} \sum_{(i,j,k)\in S} \delta_{ijk} x_{ijk}.$$

Since

$$\sum_i \delta_{ijk} = \sum_j \delta_{ijk} = \sum_k \delta_{ijk} = 1,$$

from (5.76), (5.77) and (5.80) ,

$$\begin{aligned} x_{i\cdot\cdot} &= \mu + \alpha_i^A + \nu^{-1} \sum_i \sum_j \delta_{ijk}(\alpha_{ij}^{AB} + \alpha_{ijk}^{ABC} + e_{ijk}) , \\ x_{\cdot j\cdot} &= \mu + \alpha_j^B + \nu^{-1} \sum_i \sum_k \delta_{ijk}(\alpha_{ik}^{AC} + \alpha_{ijk}^{ABC} + e_{ijk}) , \\ x_{\cdot\cdot k} &= \mu + \alpha_k^C + \nu^{-1} \sum_i \sum_j \delta_{ijk}(\alpha_{ij}^{AB} + \alpha_{ijk}^{ABC} + e_{ijk}) . \end{aligned} \qquad (5.81)$$

Lemma 5.5.

(i) $E(\delta_{ijk}) = \nu^{-1}$,

(ii)

$$E(\delta_{ijk}\delta_{i'j'k'}) = \begin{cases} \nu^{-1}, & \text{if } i = i',\ j = j',\ k = k' , \\ 0, & \text{if exactly two of the 3 equalities } i = i',\ j = j',\ k = k' \text{ hold,} \\ \nu^{-1}(\nu-1)^{-1}, & \text{if exactly one of the 3 equalities } i = i',\ j = j',\ k = k' \text{ holds,} \\ \nu^{-1}(\nu-1)^{-2}(\nu-2), & \text{if } i \neq i',\ j \neq j',\ k \neq k' . \end{cases}$$

Proof. (i) Since the randomization is equivalent to randomizing on rows, columns and symbols (treatments) so that all symbols k have the same probability of appearing in the (i,j) cell (ith row and the jth column),

$$E(\delta_{ijk}) = P\{\delta_{ijk}=1\} = \nu^{-1} .$$

(ii) If $i = i'$, $j = j'$, $k = k'$, by Lemma 5.2,

$$E(\delta_{ijk}^2) = P(\delta_{ijk}=1) = \nu^{-1}$$

For $k \neq k'$,

$$E(\delta_{ijk}\delta_{ijk'}) = P\{\delta_{ijk}\delta_{ijk'} = 1\}$$

$$= P\{\delta_{ijk} = 1\} P\{\delta_{ijk'} = 1 | \delta_{ijk} = 1\}$$

$$= \nu^{-1} P\{\delta_{ijk'} = 1|\delta_{ijk} = 1\} = 0 ,$$

as the conditional probability that the symbol k' falls in the (i,j)th cell given that the symbol k falls in the (i,j,k)th cell is zero. Thus from the symmetry of the latin square design with respect to the three factors A,B,C,

$E(\delta_{ijk}\,\delta_{i'j'k'}) = 0$ if exactly two of the 3 equalities $i = i'$, $j = j'$, $k = k'$ holds.

For $i = i'$, $j \neq j'$, $k \neq k'$, consider the symbols in the ith row only when some permutation of the symbols $1,\ldots,\nu$ appear. Since the randomization generates a random permutation of the symbols all $\nu!$ permutations are equally likely in the ith row. Of these permutations $(\nu-1)!$ have the symbol k in the jth column and of these $(\nu-2)!$ have k' in the j'th column. Thus the conditional probability

$$P\{\delta_{ij'k'} = 1|\delta_{ijk} = 1\} = \frac{(\nu-2)!}{(\nu-1)!} = \frac{1}{\nu-1} .$$

Hence

$$E(\delta_{ijk}\delta_{ij'k'}) = \nu^{-1}(\nu-1)^{-1} .$$

For $i \neq i'$, $j \neq j'$, $k \neq k'$ the conditional probability that the symbol k will be in the (i',j') cell given that the symbol k in the (i,j) cell will be the same for all $k' \neq k$, because of the random permutation of the symbols. Hence

$$(\nu-1)^{-1}(1-P\{\text{ k is in the } (i',j') \text{ cell} | \text{ k is in the } (i,j) \text{ cell}\})$$

$$= (\nu-1)^{-1}(1-P\{\delta_{i'j'k} = 1|\delta_{ijk} = 1\}) = (\nu-1)^{-1}(1-(\nu-1)^{-1}) .$$ So

$$E(\delta_{ijk}\delta_{i'j'k'}) = \nu^{-1}(\nu-1)^{-2}(\nu-2) .$$

Q.E.D.

Using the notations of Table 5.8 and Lemma 5.5, one can, after some straightforward but complicated computations, obtain from (5.80) the following theorem.

Theorem 5.6.

$$E\left(\frac{Q_3}{\nu-1}\right) = \sigma_e^2 + (1-2\nu^{-1})\ \sigma_{ABC}^2 + \sigma_{AB}^2 + \nu\sigma_C^2 ,$$

$$E\left(\frac{Q_2}{\nu-1}\right) = \sigma_e^2 + (1-2\nu^{-1})\ \sigma_{ABC}^2 + \sigma_{BC}^2 + \nu\sigma_B^2 ,$$

$$E\left(\frac{Q_1}{\nu-1}\right) = \sigma_e^2 + (1-2\nu^{-1})\ \sigma_{ABC}^2 + \sigma_{BC}^2 + \nu^2\sigma_A^2 ,$$

$$E\left(\frac{Q_4}{\nu^2-3\nu+2}\right) = \sigma_e^2 + (1-3\nu^{-1})\ \sigma_{ABC}^2 + \sigma_{AB}^2 + \sigma_{AC}^2 + \sigma_{BC}^2 ,$$

where

$$\sigma_A^2 = (\nu-1)^{-1} \sum_i (\alpha_i^A)^2 ,$$

$$\sigma_{AB}^2 = (\nu-1)^{-2} \sum_i \sum_j (\alpha_{ij}^{AB})^2 ,$$

$$\sigma_{ABC}^2 = (\nu-1)^{-3} \sum_i \sum_j \sum_k (\alpha_{ijk}^{ABC})^2 ,$$

and similar expression for other quantities like σ_B^2, σ_{BC}^2, etc. From Theorem 5.6 we may conclude under the hypothesis H_C (or equivalently H_T in Section (5.3.3)) all $\alpha_C^k = 0$ or $\sigma_C^2 = 0$, that is

$$E\left(\frac{Q_4}{\nu^2-3\nu+2}\right) > E\left(\frac{Q_3}{\nu-1}\right) \tag{5.82}$$

if $\sigma_{ABC}^2 < \nu(\sigma_{AC}^2 + \sigma_{BC}^2)$.

From (5.77), (5.81), and Lemma (5.5)

$$E(X_{..k}) = \mu + \alpha_k^C$$

no matter what interactions are present. Hence

$$\hat{\psi} = \sum_k c_k X_{..k} \text{ is an unbiased estimate of the contrast}$$

$$\psi = \sum_k c_k \alpha_k^C \quad \text{with} \quad \sum_k \alpha_k^C = 0 .$$

The calculation of $\mathrm{Var}(\hat{\psi})$, using (5.81) and Lemma (5.5), is straightforward but complicated and nothing is known about its estimate. However, from Theorem (5.6) and (5.81) if $\sigma^2_{ABC} = 0$, it can be argued for the estimates $x_{..k}$ of $\mu + \alpha^C_k$ that the errors

$$x_{..k} - \mu - \alpha^C_k , \quad k = 1, \ldots, \nu,$$

are symmetrically distributed, and in particular, they have equal variances and covariances. Under these assumptions we can prove the following theorem.

Theorem 5.7. If $\sigma^2_{ABC} = 0$, then

$$\mathrm{Var}(\hat{\psi}) = \nu^{-1}(\sigma^2_e + \sigma^2_{AB}) \sum_k c^2_k . \tag{5.83}$$

Proof. Write, for $k = 1, \ldots, \nu$

$$\theta_k = \mu + \alpha^C_k ,$$

$$\hat{\theta}_k = x_{..k} = \theta_k + f_k , \tag{5.84}$$

where

$$f_k = \nu^{-1} \sum_i \sum_j \delta_{ijk}(\alpha^{AB}_{ij} + e_{ijk}) .$$

Obviously

$$E(\hat{\theta}_k) = \theta_k , \quad E(f_k) = 0 .$$

Since $\sum_k C_k = 0$, $\hat{\psi} = \sum_k C_k \hat{\theta}_k$ is an unbiased estimator of ψ.

Let

$$Q = \frac{\sum_k (\hat{\theta}_k - \hat{\theta}_.)^2}{\nu - 1} , \quad \text{where} \quad \hat{\theta}_. = \nu^{-1} \sum_k \hat{\theta}_k .$$

As, with $\theta_. = \nu^{-1} \sum_k \theta_k$,

$$\hat{\theta}_k - \hat{\theta}_. = \theta_k - \theta_. + (f_k - f_.),$$

$$E(Q) = \sum_k \frac{(\theta_k - \theta_.)^2}{\nu - 1} + \frac{1}{\nu - 1} E(\sum_k (f_k - f_.)^2)$$

$$= \sum_k \frac{(\theta_k - \theta_.)^2}{\nu - 1} + \frac{1}{\nu - 1} (\sum_k E f^2_k - \nu E(f^2_.))$$

$$= \sum_k \frac{(\theta_k-\theta_{.})^2}{\nu-1} + \frac{1}{\nu-1} \left(\sum_k E(f_k^2) - \nu^{-1} E\left(\sum_k f_k \cdot \sum_{k'} f_{k'} \right)\right) . \qquad (5.85)$$

Under the assumption $\sigma^2_{ABC} = 0$, for all $k \neq k'$,

$$E(f_k^2) = \operatorname{Var}(f_k) = \sigma_1^2 \text{ (say)},$$

$$E(f_k f_{k'}) = \operatorname{Cov}(f_k, f_{k'}) = \rho\sigma_1^2 \text{ (say)}. \qquad (5.86)$$

From (5.85) and (5.86)

$$E(Q) = \sum_k \frac{(\theta_k-\theta_{.})^2}{\nu-1} + \frac{1}{\nu-1} \left(\nu\sigma_1^2 - \nu^{-1}(\nu\sigma_1^2 + \nu(\nu-1)\ \rho\sigma_1^2\ )\right)$$

$$= \sum_k \frac{(\theta_k-\theta_{.})^2}{\nu-1} + \sigma_1^2(1-\rho) . \qquad (5.87)$$

Now

$$\operatorname{Var}(\hat{\psi}) = E\left(\sum_k c_k(\hat{\theta}_k-\theta_k)\right)^2$$

$$= E\left(\sum_k c_k f_k\right)^2$$

$$= E\ \sigma_1^2 \sum_k c_k^2 + \rho\sigma_1^2 \sum_{\substack{k \\ k \neq k'}} \sum_{k'} c_k c_{k'}$$

$$= \sigma_1^2\ (1-\rho^2) \sum_k c_k^2 .$$

From (5.84) and (5.87) we get

$$\sigma_1^2(1-\rho) = \nu^{-1}(\sigma^2_{AB} + \sigma^2_e) . \qquad \text{Q.E.D.}$$

From Theorem 5.6 it is obvious that

$$\nu^{-1} \frac{q_4}{\nu^2-3\nu+2} \sum_k c_k^2$$

is an overestimate of $\operatorname{Var}(\hat{\psi})$ unless $\sigma^2_{AC} = 0$ and $\sigma^2_{BC} = 0$.

(b) Permutation tests. With the model (5.76) there does not exist a permutation test of H_C: all $\alpha_k^C = 0$ or equivalently

(5.85) $\sigma^2_C = 0$. A permutation test of the hypothesis H'_C that the factor C (treatments) has absolutely no effect in the sense that

$$\sigma^2_C = \sigma^2_{AC} = \sigma^2_{BC} = \sigma^2_{ABC} = 0$$

(5.86) and that the ν^2 technical errors e_{ijk} occurring in the experiment are distributed like ν^2 variables e_{ij} whose joint distribution does not depend on how the levels of C are combined with the levels of A and B. The joint distribution of the e_{ij} is subject to the conditions $E(e_{ij}) = 0$ but is otherwise arbitrary. Under H'_C , from (5.76),

(5.87)
$$x_{ijk} = \mu + \alpha^A_i + \alpha^B_j + \alpha^{AB}_{ij} + e_{ij} \quad . \qquad (5.88)$$

Since the particular latin square used in the experiment is selected at random from only one transformation set, the group G (as in the case of the randomized block design) of permutations is the set of all $(\nu!)^3$ permutations of the ν rows, the ν columns and the ν treatments (symbols). If the transformation set was first chosen with certain probabilities from the totality of all transformation sets the permutation distribution of any test statistic or its moments are determined by those for the transformation set by weighting with the same probabilities. The permutation test of H'_C based on the statistic (see Table 5.4 for notations)

$$F = (\nu-2) \quad \frac{Q_3}{Q_4}$$

Q.E.D.

is equivalent to the test based on the statistic.

$$U = \frac{Q_3}{Q_3+Q_4}, \qquad (5.89)$$

which is a strictly increasing function of F. Since

$$Q_3+Q_4 = \sum_{(i,j,k)\varepsilon S} x^2_{ijk} - \nu^2 x^2_{\cdot\cdot\cdot} - Q_2 - Q_3 , \qquad (5.90)$$

the denominator of U is constant for all permutations in G. Furthermore, since

$$Q_3 = \nu^{-1} \sum_k (\sum_i \sum_j x_{ijk})^2 - (\sum_{(i,j,k)\varepsilon S} x_{ijk})^2$$

and $\sum_{(i,j,k)\varepsilon S} X_{ijk}$ is constant for all permutations in G, it would be simplest to base the permutation test of H_C' on the statistic

$$\sum_k T_k^2 \ , \quad T_k = \sum_i \sum_k X_{ijk} \ .$$

Now, since the number of different squares in a transformation set in $\nu!(\nu-1)!$ times the number of standard squares in the set and the statistic U is invariant under the $\nu!$ permutations of ν symbols (levels of C), we may conclude that the number of different values the statistic U takes on with equal probability for a given transformation set is $(\nu-1)!$ times the number of standard squares in the set. A systematic way of calculating these values is to arrange first the ν^2 observations in a square with the rows and the columns corresponding to the levels of A, B respectively and then applying for each of the standard squares of the transformation set the $(\nu-1)!$ squares which are obtainable by permuting the $(\nu-1)$ rows of the standard squares after the first. One could also very well calculate the value of the statistic U first only for the latin squares that look as though they might give the statistic U a value greater than or equal to that observed. In other words one calculates the probability that U is greater than the observed value.

From Theorem 5.6, it is easy to see that E(U) is the same in the permutation distribution as in the distribution of U under the assumption of normality of the X_{ijk}. Welch (1937) has calculated Var(U) and the expression is very complicated for practical use. Thus the corrections needed in the degrees of freedom (as in the case of the randomized block design) of the usual test based on the assumption of normality of the X_{ijk} to make it approximate the exact permutation test better is hence not feasible to calculate. However Welch (1937) on the basis of four numerical examples showed that the normal theory test based on U is not too bad an approximation to the permutation test if $\nu > 4$.

5.3.6. Efficiency and the Optimality of the Latin Square Design:

In a latin square design it is desirable to find out if the classification of the experimental units by rows or by columns or by both has led to increased precision in the experiment. We shall find out here the efficiency of the latin square designs relative to the randomized block design and to the completely randomized design. The computation in this case goes in the same way as in the case of the randomized block design. So we shall merely state the results.

The estimated efficiency of the latin square design relative to the randomized block design made by rows as blocks, is given by

$$\frac{\dfrac{q_2}{\nu-1} + (\nu-1)\dfrac{q_4}{(\nu-1)(\nu-2)}}{\dfrac{\nu q_2}{(\nu-1)(\nu-2)}} = \frac{(\nu-2)\ q_2 + (\nu-1)\ q_4}{\nu q_4}$$

The estimated efficiency of the latin square design relative to the randomized block design made up by columns as blocks, is given by

$$\frac{(\nu-2)q_1 + (\nu-1)q_4}{\nu q_4}$$

The efficiency of the latin square design relative to the completely randomized design is given by,

$$\frac{\dfrac{(q_1+q_2)}{\nu-1} + (\nu-1)\dfrac{q_4}{(\nu-1)(\nu-2)}}{(\nu+1)\dfrac{q_4}{(\nu-1)(\nu-2)}} = \frac{(\nu-2)(q_1+q_2)+(\nu-1)q_4}{(\nu+1)q_4}$$

The quantities q_1, q_2, q_4 are given in Table 5.3.

Cochran (1938, 1940) reported that the efficiency of the latin square design relative to the completely randomized design is 222 percent and relative to the randomized block design is 137 percent for the experiments conducted at Rothamstead and its associated centres during 1927-1934. A similar result was obtained by Ma and Harrington (1948) who found that the completely randomized design was 79 percent as efficient as the latin square design.

However, it may be remarked that if a randomized block design has to be used instead of a latin square design, the best possible shape of a block from the point of view of the experimental procedure might have to be of a shape other than that given by rows or columns. This may vitiate many comparisons that are made in this way.

The fact that the latin square design is A-, D-, E- and L-optimum will follow from the A-, D-, E- and L-optimum properties of the generalized Youden square design which we shall discuss in Section 5.4.

5.4. BALANCED INCOMPLETE BLOCK DESIGNS

If the number of treatments ν in an experiment is large, it may not be possible for various reasons to use large size blocks to accommodate all treatments in each block.

We may use incomplete block designs, that is, designs in which the number of experimental units in each block is less than the total number of treatments.

Let us consider an experiment where ν treatments are to be compared in b blocks of sizes $k_1,\ldots,k_b$. Let the ith treatment be replicated in r_i blocks, $i = 1,\ldots,\nu$ and let the ith and the jth treatment occur together in λ_{ij} blocks, $i,j = 1,\ldots,\nu$, $i \neq j$.

The $b \times \nu$ matrix $N = (n_{ij})$, where n_{ij} is the number of times the jth treatment occurs in the ith block is called the incident matrix of the incomplete block design. If n_{ij} takes only values 1 or 0 then the incomplete block design is called a binary design.

The matrix

$$C = D(r_1,\ldots,r_\nu) - N'D(k_1^{-1},\ldots,k_b^{-1})N\ , \tag{5.91}$$

where $D(\theta_1,\ldots,\theta_k)$ is a $k \times k$ diagonal matrix with diagonal elements $\theta_1,\ldots,\theta_k$, is called the C-matrix of the incomplete block design.

Definition 5.1. A contrast $\sum_j c_j t_j = \underline{c}'\underline{t}$, with $\underline{c} = (c_1,\ldots,c_\nu)'$ in the treatment effects $\underline{t} = (t_1,\ldots,t_\nu)'$ is called an elementary contract if $\underline{c}$ has only two non-zero components 1 and -1.

For the proof of various results presented below in connection with the properties of incomplete block design we refer to Raghava Rao (1971).

Elementary contrasts in the treatment effects involve all differences of $t_i - t_j$, $i \neq j$ and it is often desirable to design experiments where all elementary contrasts are estimable.

Definition 5.2. Connected design. A design where all elementary contrasts are estimable is called a connected design. Otherwise it is called a disconnected design.

An incomplete block design with ν treatments is connected if and only if the rank of its C-matrix is ν-1.

Chakrabarti (1963) has shown that in a connected design all the diagonal elements of its C-matrix are positive and the principal minors of all orders of its C-matrix are positive.

Definition 5.3. Balanced design. A connected design is said to be balanced if all elementary contrasts in the treatment effects can be estimated with the same precision (inverse of the variance of the estimator).

This definition does not hold for the disconnected design, as all elementary contrasts are not estimable in this design. To overcome this difficulty, Vartak (1963) defined a design (not necessarily connected) to be balanced if every estimable normalized contrast in the treatment effects can be estimated with the same precision. Rao (1958) obtained the following two theorems.

Theorem 5.8. A connected design is balanced if and only if all characteristic roots of its C-matrix are equal.

Let the nonzero characteristic roots of C be all equal to θ. Then each of the elementary contrasts, in the treatment effects is estimated with variance $2\theta^{-1}\sigma^2$ where σ^2 is the error variance and the design is balanced. For the balanced design the C-matrix is of the form

$$C = \theta(I_\nu - \nu^{-1}E_{\nu\nu}) \tag{5.92}$$

where I_ν is the $\nu \times \nu$ identity matrix and $E_{\nu\nu}$ is the $\nu \times \nu$ matrix whose elements are all equal to unity.

Theorem 5.9. For any balanced design with $r_1 = \dots = r_\nu = r$ (say), $k_1 = \dots = k_b = k$ (say), the incident matrix N satisfies

$$N'N = (r-\nu)\, I_\nu + E_{\nu\nu} \tag{5.93}$$

where $\lambda(\nu-1) = r(k-1)$.

Note. An incomplete block design with $k_1 = \dots = k_b = k$ is called a proper design.

Definition 5.4. Balanced incomplete block designs (BIB). A balanced incomplete block design is an arrangement of ν treatments (symbols) in b blocks each containing k experimental units $(k < \nu)$ such that

(i) every treatment occurs at most once in each block,
(ii) every treatment occurs exactly in r blocks,
(iii) every pair of treatments occur together in exactly λ blocks.

The quantities ν, b, r, k, λ are called the parameters of the BIB design.

From Definition 5.4 it follows that

$$\nu r = bk \tag{5.94}$$

The incident matrix N of the BIB design satisfies Theorem 5.9 and the characteristic roots of its C-matrix are equal. Thus from Theorem 5.8 it is a balanced design.

Example 5.6.

Block	Treatments	Block	Treatments
1	1, 2, 4, 7	5	5, 6, 1, 4
2	2, 3, 5, 1	6	6, 7, 2, 5
3	3, 4, 6, 2	7	7, 1, 3, 6
4	4, 5, 7, 3		

is an example of a BIB design with parameters $\nu = b = 7$, $r = k = 4$, $\lambda = 2$.

Although Kirkman (1847) found a particular series of BIB designs, its systematic development in statistical literature was originated by Yates (1936). Since then significant contributions were made, among others, by Fisher (1940, 1942), Fisher and Yates (1963) and Bose during the years 1938 to 1959. Though the following lemma follows from Theorem 5.9, it is worth giving the following elementary proof of it here.

Lemma 5.10. For a BIB design with parameters ν, b, r, k,

$$\lambda(\nu-1) = r(k-1). \tag{5.95}$$

Proof. Consider all possible pairs of treatments each containing a particular treatment θ. Since θ occurs in r blocks in the design and in each of these r blocks there are k-1 other treatments present, the total number of possible pairs of treatments in the design, each containing treatment θ is $r(k-1)$.

On the other hand from Definition 5.4 of the BIB design, θ is paired with each of the remaining $\nu-1$ treatments exactly λ times. So the total number of possible pairs of treatments each containing θ is $\lambda(\nu-1)$. Hence

$$\lambda(\nu-1) = r(k-1)$$

Q.E.D.

From Theorem 5.9 the incident matrix N of the BIB design satisfies $N'N = (r-\lambda)I_\nu + E_{\nu\nu}$. Thus

$$|N'N| = (r + (\nu-1)\lambda)(r-\nu)^{\nu-1}. \tag{5.96}$$

Since $k < \nu$ from Lemma 5.10 we get $\lambda < r$. Thus from (5.96), N'N is a nonsingular matrix. Now

$$\nu = \text{rank}(N'N) = \text{rank}(N) \leq b. \tag{5.97}$$

This is due to Fisher (1940).

Definition 5.5. A BIB design which can be arranged in sets of blocks such that each set contains a complete replicate of all the treatments is called resolvable.

In the case of resolvable BIB designs Bose (1942) showed that the inequality (5.97) can be improved to

$$\nu \leq b-r+1 \tag{5.98}$$

If for a resolvable BIB design the equality in (5.98) holds, then it is called affine resolvable. For the existence and the construction of BIB designs for different values of the parameters we refer to Raghava Rao (1971).

Generalizing the notion of BIB designs Kiefer (1958) defined a design to be a balanced block design (BBD) if it is an arrangement of ν treatments in b blocks each containing k experimental units (not necessarily $k \leq \nu$) such that

(i) the number of times n_{ij} that the treatment j appears in the ith block is k/ν if this is an integer and is one of the two closest integers otherwise,

(ii) the number of replications of the treatment j is the same for all j,

(iii) for each pair of treatments (j,j'), $j \neq j'$,

$\sum_i n_{ij} n_{ij'}$ is the same.

Certain designs of this type were constructed by Tocher (1952). A special case of BBD, where each of the n_{ij} takes one of three preassigned values, was considered by Rao and Das (1969), which they called balanced ternary design (BTD).

5.4.1. Intra-block Analysis of BIB Designs: We shall consider here the intra-block analysis of a BIB design with parameters ν,b,r,k,λ under the normal theory fixed effect model. It is slightly more complicated than that of the other complete block designs treated so far. Let x_{ij} denote the observation on the jth treatment in the ith block (provided that the jth treatment occurs in the ith block) where (i,j) runs in the set S for which $n_{ij} = 1$. Let us assume that the corresponding random variables X_{ij} are independently normally distributed with

$$E(X_{ij}) = \mu + \alpha_i + t_j$$

$$\mathrm{Var}(X_{ij}) = \sigma^2 \qquad (5.99)$$

or

$$x_{ij} = \mu + \alpha_i + t_j + e_{ij}$$

$$(5.100)$$

where $(i,j) \varepsilon S$, the e_{ij} are independently normally distributed with mean 0 and variance σ^2, μ is the grand mean, the α_i is the ith block effect satisfying $\sum_i \alpha_i = 0$ and the t_j is the jth treatment effect satisfying $\sum_j t_j = 0$. We are interested in testing the hypothesis H_T: all $t_j = 0$. The hypothesis H_α: all $\alpha_i = 0$, which has no practical significance can be treated similarly. The problem of testing H_T is a simple case of a two-way classification with unequal number of observations per cell and without any interactions (as treated in Chapter 3).

First we shall consider the analysis of a general incomplete block design (not balanced) with the added restriction that $r_1 = \ldots = r_\nu = r$ and $k_1 = \ldots = k_b = k$. Then we shall specialize for the case of a BIB design. In the notation of Section 3.1, the hypothesis H_T is a linear hypothesis with $r = \nu-1$, $s = \nu+b-1$, $N-s = bk - \nu - b+1 = \nu r - \nu - b+1$. The least squares estimates of μ, α_i, t_j under Ω can be obtained by minimizing

$$s^2 = \sum_{(i,j) \varepsilon S} (x_{ij} - \mu - \alpha_i - t_j)^2 \tag{5.101}$$

with respect to μ, α_i and t_j. Equating to zero the partial derivatives of s^2 with respect to μ, α_i and t_j, we get for $i = 1,\ldots,b;\ j = 1,\ldots,\nu$,

$$\begin{aligned} bk\,\hat{\mu} + k \sum_i \hat{\alpha}_i + r \sum_j \hat{t}_j &= \sum_{(i,j) \varepsilon S} x_{ij}, \\ k(\hat{\mu} + \hat{\alpha}_i) + \sum_j n_{ij}\,\hat{t}_j &= B_i, \\ r(\hat{\mu} + \hat{t}_j) + \sum_i n_{ij}\,\hat{\alpha}_i &= T_j \end{aligned} \tag{5.102}$$

where

B_i = ith block total = $\sum_j x_{ij}$,

T_j = jth treatment total = $\sum_i x_{ij}$,

$r = \sum_i n_{ij}$, $k = \sum_j n_{ij}$.

Let

$$\sum_i n_{ij}\, n_{ij'} = \lambda_{jj'}\,, \tag{5.103}$$

$$Q_j = T_j - k^{-1} \sum_i n_{ij}\, B_i = T_j - k^{-1} T'_j\,,$$

where T'_j is the sum of all block totals in which the jth treatment occurs.

For the BIB design

$$\lambda_{jj'} = \begin{cases} \lambda, & \text{if } j \neq j' \\ r, & \text{if } j = j' \end{cases} \,. \tag{5.104}$$

The quantity Q_j is called the jth adjusted treatment total, the adjustment consists in subtracting from the jth treatment total T_j the sum of block averages $k^{-1}B_i$ for all blocks in which the jth treatment occurs. From (5.102) and (5.103) the estimates of t_j, which for reasons that will be obvious later are called intra-block estimates of the t_j are given by (with $\delta_{jj'} = 1$ if $j = j'$ and 0 otherwise).

$$\sum_{j'} (r\, \delta_{jj'} - k^{-1} \lambda_{jj'})\, \hat{t}_{j'} = Q_j\,. \tag{5.105}$$

The error sum of squares for testing H_T or H_α is the minimum value of s^2 in (5.101) and is given by

$$q_3 = \sum_S x_{ij}^2 - \sum_j \hat{t}_j \hat{Q}_j - k^{-1} \sum_i B_i^2$$

$$= [\sum_S x_{ij}^2 - \text{C.F.}] - \sum_j \hat{t}_j Q_j - [k^{-1} \sum_i B_i^2 - \text{C.F.}]$$

= [Total sum of squares] - [treatment sum of squares eliminating blocks] - [block sum of squares ignoring treatments],

where $\text{C.F.} = b^{-1}k^{-1}(\sum_S x_{ij})^2$. (5.106)

Under H_T: all $t_j = 0$ the minimum value of s^2 is given by

$$\sum_S x_{ij}^2 - k^{-1} \sum_i B_i^2 \tag{5.107}$$

= [Total sum of squares] - [block sum of squares ignoring treatments].

Thus the best invariant test of level α of H_T rejects H_T whenever

$$\frac{bk-b-\nu+1}{\nu-1}\sum_j \frac{\hat{t}_j Q_j}{q_3} \geq F_{1-\alpha,\nu-1,bk-\nu-b+1} \, . \qquad (5.108)$$

We now specialize for the case of the BIB design where $\lambda_{jj'}$ satisfies (5.104). From (5.105)

$$\frac{r(k-1)}{k}\hat{t}_j - k^{-1}\lambda \sum_{j'\neq j}\hat{t}_{j'} = Q_j \qquad (5.109)$$

Because of the side condition

$$\sum_j \hat{t}_j = 0 \qquad (5.110)$$

we get from (5.109)

$$\hat{t}_j = (r\varepsilon)^{-1} Q_j \qquad (5.111)$$

where

$$\varepsilon = \frac{rk-r+\lambda}{rk} = \frac{(k-1)\nu}{k(\nu-1)} < 1 \, . \qquad (5.112)$$

If there are more blocks that contain the treatment j than do not, that is if $r > b-r$, it is easier to compute $\tilde{T}_j$, where $\tilde{T}_j$ = sum of all block totals not containing the treatment j and to use the differences of the

$$\tilde{Q}_j = \frac{T_j + k^{-1}\tilde{T}_j}{r\varepsilon} \qquad (5.113)$$

for estimating elementary contrasts in treatment effects. This is because $T_j + \tilde{T}_j = \sum_S x_{ij}$. From (5.111)

$$\sum_j \tilde{t}_j Q_j = (rE)^{-1}\sum_j Q_j^2 \, . \qquad (5.114)$$

Thus the intra-block analysis of variance table for the BIB design is given in Table 5.9.

If the F-value is significant and consequently H_T is rejected it is then necessary, by methods of multiple comparisons, to divide the entire set of treatments into groups such that the treatments of the same group have the same treatment effects and those from different groups have different treatments effects. For this purpose we need to find the common variance $\sigma^2_{\hat{t}}$ and the common variance $\rho\sigma^2_{\hat{t}}$ (ρ is the common correlation coefficient) of the intra-block estimators $\hat{t}_j$ of t_j. This can be obtained by the following simple device.

Table 5.9. Intra-block ANOVA for the BIB design

Sources of variation	d.f.	sum of squares	Expected mean squares	F
Blocks ignoring treatments	$b-1$	$q_1 = k^{-1} \sum_i B_i^2 - \text{C.F.}$		
Treatments eliminating blocks	$\nu-1$	$q_2 = (r\varepsilon)^{-1} \sum_j Q_j^2$	$\sigma^2 + r\varepsilon\sigma^2_T$	$\frac{bk-b-\nu+1}{(\nu-1)} \frac{q_2}{q_3}$
Error (intra-block)	$bk-b-\nu+1$	q_3	σ^2	
Total	$bk-1$	$\sum_S x_{ij}^2 - \text{C.F.}$		

where $\sigma_T^2 = (\nu-1)^{-1} \sum_j t_j^2$.

Note. Because of the symmetrical way in which the treatments appear in the design the $\hat{t}_j$ have equal variances and covariances. Since $\sum_j \hat{t}_j = 0$, we get

$$\nu\sigma^2_{\hat{t}} + \nu(\nu-1)\,\rho\,\sigma^2_{\hat{t}} = 0 \ , \tag{5.115}$$

and hence $\rho = -(\nu-1)^{-1}$. From the Section 5.4.3 it will follow that

$$\mathrm{Var}(\hat{t}_1 - \hat{t}_2) = (r\varepsilon)^{-1}\, 2\sigma^2 . \tag{5.116}$$

Thus from (5.115) and (5.116) we get

$$(rE)^{-1}\sigma^2 = 2\sigma^2_{\hat{t}} - 2\rho\,\sigma^2_{\hat{t}} = 2\nu(\nu-1)^{-1}\,\sigma^2_{\hat{t}}$$

and consequently

$$\mathrm{Var}(\hat{t}_j{}') = \frac{\sigma^2(\nu-1)}{r\varepsilon}$$

$$\mathrm{Cov}(\hat{t}_j, \hat{t}_{j'}) = -\frac{\sigma^2}{r\varepsilon} \ . \tag{5.117}$$

5.4.2. Missing Observations in the BIB Design: Cornish (1940) has given a formula to obtain the intra-block estimate of a missing value in the BIB design. The intra-block estimate of the missing value x_{ij} is

$$\hat{x}_{ij} = \frac{\nu r(k-1)B_i - k(\nu-1)Q_j - (\nu-1)Q'_j}{k(k-1)(bk-b-\nu+1)} \qquad (5.118)$$

where Q'_j is the sum of the Q values for all other treatment (but not the jth one) which are present in the ith block. If more than one observation is missing the same procedure as discussed in the case of the randomized block design and the latin square design may be adopted. As usual one degree of freedom per missing datum is deducted from the error sum of squares of the intra-block analysis of variance. Also one may introduce concomitant variables composed of zeros and ones for the missing values and perform the analysis of covariance (Rao, 1947). If several observations are missing it may be advisable to analyse the design as a completely randomized design.

5.4.3. The Efficiency and the Optimality of the BIB Design:

From Theorem 5.7 we can conclude that for any estimable contrast $\psi = \sum_j c_j t_j$, the variance of its unbiased estimator $\hat{\psi} = \sum_j c_j \hat{t}_j$ is given by

$$\mathrm{Var}(\hat{\psi}) = V \sum_j c_j^2$$

where

$$\begin{aligned} V &= E[(\nu-1)^{-1} \sum_j (\hat{t}_j - \hat{t}_.)^2] - (\nu-1)^{-1} \sum_j (t_j - t_.)^2 \\ &= E[(r\varepsilon)^{-1} (\nu-1)^{-1} Q'_2] - \sigma^2_T \\ &\approx (r\varepsilon)^{-1} \sigma^2 , \end{aligned}$$

where Q'_2 is the random variable corresponding to q_2 in Table 5.9. Thus

$$\mathrm{Var}(\hat{\psi}) = (t\varepsilon)^{-1} \sigma^2 \sum_j c_j^2 \qquad (5.119)$$

If we have conducted the experiment in the randomized block design with the same number of treatments ν, the same number of replications r for each treatment and the same σ^2, then the unbiased estimator of ψ is

$$\hat{\psi} = \sum_j c_j x_{.j}$$

with

$$\mathrm{Var}(\tilde{\psi}) = \frac{\sigma^2}{r} \sum_j c_j^2 \quad . \tag{5.120}$$

So the efficiency of the BIB design relative to the randomized block design is

$$\frac{\mathrm{Var}(\tilde{\psi})}{\mathrm{Var}(\hat{\psi})} = \varepsilon \tag{5.121}$$

Thus the quantity ε , as given in (5.112) is called the efficiency factor of the BIB design.

To facilitate the derivation of optimality results we need to develop the analysis of a general incomplete block design with $r_j = r$ and all $k_i = k$ (which is not necessarily a binary design). Let

$$G = \sum_{(i,j)\in S} x_{ij},\ \underline{T} = (T_1,\ldots,T_\nu)',\ \underline{B} = (B_1,\ldots,B_b)'$$

$$Q = (Q_1,\ldots,Q_\nu)',\quad \underline{t} = (t_1,\ldots,t_\nu)',\quad \underline{\alpha} = (\alpha_1,\ldots,\alpha_b)',$$

$I(n)$ = n × n identity matrix,

θ_n = n-vector whose components are all equal to unity.

From (5.102) we get

$$\begin{pmatrix} G \\ \underline{T} \\ \underline{B} \end{pmatrix} = \begin{pmatrix} bk & r\theta'_\nu & k\theta'_b \\ r\theta'_\nu & rI(\nu) & N' \\ k\theta_b & N & kI(b) \end{pmatrix} \begin{pmatrix} \hat{\underline{\mu}} \\ \hat{\underline{t}} \\ \hat{\underline{\alpha}} \end{pmatrix} \tag{5.122}$$

Since

$$r\theta_\nu = N'\theta_b\ ,\quad k\theta_b = N\theta_\nu\ , \tag{5.123}$$

multiplying both sides of (5.122) by

$$\begin{pmatrix} n^{-1}k^{-1} & 0 & 0 \\ 0 & I(\nu) & -k^{-1}N' \\ 0 & -r^{-1}N & I(b) \end{pmatrix},$$

we obtain

$$\begin{aligned} \underline{Q} &= (rI(\nu) - k^{-1}N'N)\underline{\hat{t}} \\ &= C\,\underline{\hat{t}} \end{aligned} \qquad (5.124)$$

where

$C = rI(\nu) - k^{-1}N'N$,

N is the incident matrix

Using (5.123) we get

$$C\theta_\nu = \theta_\nu' \; C = 0 \qquad (5.125)$$

So we confine our discussion to the case where the matrix C has rank $\nu-1$. Thus this design is connected and all contrasts $t_j - t_{j'}$, $(j \neq j')$ are estimable. It is now left as an exercise to show that (Note: here $\underline{Q}$ is a random vector)

$\operatorname{Var}(\underline{Q}) = C\sigma^2$.

Since for any general incomplete block design

$\sum_{i,j} n_{ij} = bk,$

$\operatorname{tr} C = bk - k^{-1} \sum_{i,j} n_{ij}^2$ is maximized by the balanced block design whenever it exists. The C matrix of such a design is of the form hypothesized in (5.88). Thus the balanced block design is A-, D-, E- and L-optimum. In particular the randomized block design and the BIB design are A-, D-, E- and L-optimum. This proof is due to Kiefer (1958).

For binary designs Mote (1958) proved the E-optimality of BIB designs and Kshirsagar (1958) proved the A- and the D-optimality of BIB designs for estimating orthogonal contrasts and also of the Youden square design. Roy (1958) established the necessary and the sufficient condition for a binary design to be most efficient with respect to the A-optimality criterion for estimating all elementary contrasts and proved that in the case of proper binary incomplete block designs a most efficient design, if it exists, is necessarily a balanced incomplete block design.

5.4.4. Recovery of Inter-block Information: Since some incomplete block designs have low efficiency factors ε , Yates (1939, 1940) suggested the use of informations available from interblock comparisons to increase the precision of estimates of treatment effects and showed how this can be done for the cubic lattice design (1939) (to be discussed later in Chapter 6) and BIB design (1940). Nair (1944) gave the method for a partially balanced incomplete block (PBIB) design (to be discussed later in this chapter). Giri (1957) gave the method for the row balanced PBIB design. Rao (1947, 1956) adopted this method for any incomplete block design. Later Roy and Shah (1962) and Shah (1964) examined the problem of recovery of interblock information in detail in the case of a general incomplete block design.

The process of recovery of interblock information consists of applying the method of weighted least squares to intrablock and interblock contrasts of observations for the purpose of estimating the treatment effects, the weights being inversely proportional to the variances of these two contrasts. The ratio ρ of the interblock variance to the intrablock variance plays an important role and since this is usually unknown the ratio of estimates of these variances obtained from an analysis of variance of the observed data is substituted (Yates (1939, 1940)). If ρ were known, the combined estimator would have all the good properties of the usual least squares estimators. Since only an estimate of ρ is used, obviously the properties of this combined estimator are dependent on the type of estimator of ρ used. Graybill and Weeks (1959) showed that for a BIB design the combined estimator of a treatment contrast based on Yates' estimator of ρ in its untruncated form is unbiased. Graybill and Seshadri (1960) proved the same result with estimator of ρ in its usual truncated form for BIB design. Shah (1969) has obtained an estimator of ρ which produces a combined estimator which is uniformly better than the intrablock estimator for any treatment contrast for a BIB design with more than 5 treatments, for a simple lattice design with 16 treatments or more and for a triple lattice design with 9 treatments or more.

Now we shall treat here in detail the case of a BIB design. It can be easily extended to Youden square designs, the PBIB designs and the row balanced PBIB designs. It may be remarked that if we stick to the fixed-effect model of the BIB design as given in (5.100), then intrablock estimators are optimum by the Gauss-Markoff theorem (Chapter 3). The problem of recovery of interblock information to obtain more efficient estimators of treatment contrasts arises if we consider the block effects α_i in (5.100) to be random variables a_i instead of unknown constants α_i satisfying $\sum_i \alpha_i = 0$.

The new mathematical model which we shall now adopt for the recovery of interblock informations is not usually realistic for field experimentations where the a_i are random variables only because they are random permutations of the fixed block effects of the b blocks actually used in the experiment, in this case a randomization model already discussed in 5.2 and 5.3 would fit better. Our assumption for the recovery of interblock information is as follows.

Let

$$x_{ij}' = \mu + a_i + t_j + e_{ij}, \quad (i,j) \in S \tag{5.126}$$

where the e_{ij} are independently, normally distributed with mean 0 and variance σ^2 and are independent of the a_i. The a_i are independently normally distributed with mean 0 and the variance σ_B^2 and $\sum_j t_j = 0$. We cannot assume that $a_. = \sum_i a_i = 0$, because the a_i are independent.

We shall first observe that under the model (5.126) the intrablock estimators $\hat{t}_j$ and the test statistic $\left(\frac{bk-b-v+1}{v-1}\right)\frac{Q_2}{Q_3}$ where Q_2 and Q_3 are random variables corresponding to q_2 and q_3 respectively (as given in Table 5.9) have the same distributions as under (5.100). Since the joint distribution of the $\hat{t}_j$, Q_2 and Q_3 can be regarded as the conditional distribution under (5.126) given the a_i, the constant μ and the α_i in (5.100) corresponding to the fixed values of the a_i according to $\alpha_i = a_i - a_./b$, $\mu = \mu' + a_./b$. Since this conditional distribution does not depend on the values of μ and the α_i and hence not on the a_i, it must be same as the unconditional distribution of the $\hat{t}_j$, Q_2 and Q_3.

Let $\hat{t}_j$ denote the least squares estimates of the t_j from the block totals

$$\begin{aligned} B_i &= \sum_j n_{ij} (\mu' + a_i + t_j + e_{ij}) \\ &= k\mu' + \sum_j n_{ij} t_j + f_i, \quad i = 1,\ldots,b \end{aligned} \tag{5.127}$$

where

$$f_i = k a_i + \sum_j n_{ij} e_{ij} . \tag{5.128}$$

The f_i are independently normally distributed with mean 0 and the common variance $\sigma_f^2 = k^2 \sigma_B^2 + k\sigma^2$. Equation (5.127) represents a fixed effect linear model in terms of the random variables B_i, $i = 1,\ldots,b$, (we are using here the same notations for the random variable and its values) with

$$\begin{aligned} E(B_i) &= k\mu' + \sum_j n_{ij} t_j , \\ \mathrm{Var}(B_i) &= k^2 \sigma_B^2 + k\sigma^2, \quad i = 1,\ldots,b . \end{aligned} \tag{5.129}$$

Now minimizing

$$s^2 = \sum_i (B_i - k\mu' - \sum_j n_{ij} t_j)^2 \tag{5.130}$$

with respect to μ' and t_j we obtain

$$\begin{aligned} k^2 b\, \hat{\mu}' &= k \sum_i B_i \\ kr\hat{\mu} + \sum_{j'} \lambda_{jj'} \hat{t}_{j'} &= T'_j \end{aligned} \tag{5.131}$$

where $\lambda_{jj'} = \sum_i n_{ij} n_{ij'}$ and T'_j is the sum of all block totals containing the treatment j. From (5.104) and (5.131) we get

$$kr\,\hat{\mu}' + (r-\lambda)\hat{t}_j = T'_j \tag{5.132}$$

From (5.131) - (5.132) we get

$$\hat{\hat{t}}_j = (r-\lambda)^{-1}(T'_j - rb^{-1}\sum_i B_i) \tag{5.133}$$

Since $\sum_i B_i = \sum_S x_{ij}$, the interblock estimate of any treatment contrast

$\hat{\hat{\psi}} = \sum_j c_j t_j$ with $\sum_j c_j = 0$ is given by

$$\hat{\hat{\psi}} = (r-\lambda)^{-1} \sum_j c_j T'_j \ .$$

If 2r > b, it is easier to compute $(r-\lambda)^{-1}\tilde{T}_j$ and use $\hat{\hat{\psi}} = -(r-\lambda)^{-1}\sum c_j \tilde{T}_j$. The error sum of squares based on the observations B_i is the minimum value of s^2 in (5.130) and is given by

$$SS_f = \sum_i B_i^2 - \hat{\hat{\mu}}' k \sum_i B_i - \sum_j \hat{\hat{t}}_j T'_j$$

$$= \sum_i B_i^2 - (r-\lambda)^{-1} \sum_j (T'_j)^2 + \frac{\nu(k-1)}{b(\nu-k)} \left(\sum_i B_i\right)^2 \tag{5.134}$$

as $\sum_j T'_j = \sum_j \sum_i n_{ij} B_i = k \sum_i B_i$, $(\nu-1)\lambda = r(k-1)$. If $b > \nu$, by the general theory of Chapter 3

$$\hat{\sigma}_f^2 = \frac{SS_f}{b-\nu} \tag{5.135}$$

is distributed as $(b-\nu)^{-1}\sigma_f^2 \chi^2_{b-\nu}$ and is independent of $\hat{\hat{t}}_j$ and hence of $\hat{\hat{\psi}}$.

Theorem 5.11. Under the model (5.126), the intrablock estimator $\hat{\psi}$ and the interblock estimator $\hat{\hat{\psi}}$ of any contrast $\psi = \sum_j c_j t_j$ are independent.

Proof. We shall use wherever convenient the same notation for the random variable and its values. As the t_j and μ' are constants,

$$\text{Cov}(B_i, T_j) = \text{Cov}\Big(\sum_{j'} n_{ij'} X_{ij'}, \sum_{i'} n_{i'j} X_{i'j}\Big)$$

$$= \sum_{i'} \sum_{j'} n_{ij'} n_{i'j} \, \text{Cov}(X_{ij'}, X_{i'j})$$

$$- \sum_{i'} \sum_{j'} n_{ij'} n_{i'j} \operatorname{Cov}(b_i + e_{ij'}, b_{i'} + e_{i'j})$$

$$= \sum_{i'} \sum_{j'} n_{ij'} n_{i'j} \delta_{ii'} (\sigma_B^2 + \delta_{jj'} \sigma^2)$$

$$= n_{ij} \sum_{j'} n_{ij'} (\sigma_B^2 + \delta_{jj'} \sigma^2)$$

$$= n_{ij} (k \sigma_B^2 + \sigma^2)$$

$$= n_{ij} \frac{\sigma_f^2}{k} \tag{5.136}$$

Since the a_i and the e_{ij} are independently distributed with zero means and variances σ_B^2 and σ^2 respectively, using (5.135)

$$\operatorname{Cov}(Q_j, B_i) = \operatorname{Cov}(T_j - k^{-1} \sum_{i'} n_{i'j} B_{i'}, B_i)$$

$$= \operatorname{Cov}(T_j, B_i) - k \sum_{i'} n_{i'j} \operatorname{Cov}(B_{i'}, B_i)$$

$$= k^{-1} n_{ij} \sigma_f^2 - k \sum_{i'} n_{i'j} \operatorname{Cov}(B_{i'}, B_i)$$

$$= k^{-1} n_{ij} \sigma_f^2 - k^{-1} n_{ij} \sigma_f^2 = 0, \tag{5.137}$$

as the B_i are independently normally distributed with

$\operatorname{Cov}(B_i, B_{i'}) = \delta_{ii'} \sigma_f^2$.

Since (Q_j, B_i) have a joint multivariate normal distribution from (5.137) the Q_j and hence the $\hat{t}_j$ are independent of B_i. Thus $\hat{\hat{\psi}}$, being only a function of the $\hat{t}_j$, is independent of $\hat{\hat{\psi}}$ (function of the B_i only).

Q.E.D.

Theorem 5.12. $\operatorname{Var}(\hat{\hat{\psi}}) = \sum_j c_j^2 \frac{\sigma_f^2}{r-\lambda}$

Proof. Since $\hat{\hat{\psi}} = \sum_j \frac{T'_j}{r-\lambda}$,

$$\operatorname{Var}(\hat{\hat{\psi}}) = (r-\lambda)^{-2} \sum_j \sum_{j'} C_j C_{j'} \operatorname{Cov}(T'_j, T'_{j'}) . \tag{5.138}$$

But

$$\begin{aligned}\mathrm{Cov}(T'_j, T'_{j'}) &= \mathrm{Cov}(\sum_i n_{ij} B_i, \sum_{i'} n_{i'j'} B_{i'}) \\ &= \sum_i \sum_{i'} n_{ij} n_{i'j'} \delta_{ii'} \sigma_f^2 \\ &= \sum_i n_{ij} n_{ij'} \sigma_f^2 = \lambda_{jj'} \sigma_f^2 .\end{aligned} \tag{5.139}$$

From (5.137) and (5.138), since $\sum_j C_j = 0$ we get

$$\mathrm{Var}(\hat{\psi}) = \sum_j C_j^2 \frac{\sigma_f^2}{r-\lambda} .$$

Q.E.D.

Let

$$\omega = \frac{r\varepsilon}{\sigma^2}, \qquad \omega' = \frac{(r-\lambda)}{\sigma_f^2} \tag{5.140}$$

where $\sigma_f^2 = k^2 \sigma_B^2 + k \sigma^2$. Since for any contrast $\psi = \sum_j C_j t_j$ its intrablock and interblock estimators $\hat{\psi}$ and $\hat{\hat{\psi}}$ are statistically independent, the linear estimator

$$\psi^* = \frac{\omega\hat{\psi} + \omega'\hat{\hat{\psi}}}{\omega+\omega'} \tag{5.141}$$

is the best (minimum variance) unbiased estimator of ψ. If we know σ^2 and σ_f^2 we could obtain ψ^*. In general they are unknown. The original method of Yates (1940) is to use ψ^* but with the weights obtained by estimating σ^2 and σ_f^2 from the data. The usual estimate of ω is obtained from the error of the BIB design (Table 5.9) and is given by

$$\hat{\omega} = \frac{(bk-b-\nu+1)\,r}{q_3} \tag{5.142}$$

Though one can estimate ω' by (see (5.135))

$$\hat{\omega}' = \frac{(r-\lambda)(b-\nu)}{SS_f}$$

The usual choice of ω' (Yates (1940)) is given by

$$\hat{\omega}' = \frac{r-\lambda}{\hat{\hat{\sigma}}_f^2} ,$$

where $\hat{\hat{\sigma}}_f^2$ is an unbiased estimate of σ_f^2 obtained through the mean sum of squares due to blocks eliminating treatments. This is obtained from the following analysis of variance table for testing H_α: all $\alpha_i = 0$ (see Table 5.9 for notation).

Table 5.10. Test of H_α for BIB design

Sources of Variation	d.f.	Sum of squares	F
Treatments (ignoring blocks)	$\nu-1$	$u_1 = r^{-1} \sum_j T_j^2 - C.F.$	
Blocks (eliminating treatments)	$b-1$	u_2 (by subtraction)	$\frac{bk-b-\nu+1}{b-1} \frac{u_2}{q_3}$
Error (intrablock)	$bk-b-\nu+1$	q_2	
Total	$bk-1$	$\sum_{(i,j)\varepsilon S} x_{ij}^2 - C.F.$	

so the sum of squares for blocks eliminating treatments is given by

$$SS(bl.adj) = k \sum_i B_i^2 + (r\varepsilon)^{-1} \sum_j Q_j^2 - r^{-1} \sum_j T_j^2 \tag{5.143}$$

Theorem 5.13. Under the model (5.100)

$$E(SS(bl.adj) = (b-1)\sigma^2 + (bk-\nu)\sigma_B^2 . \tag{5.144}$$

Note. In what follows we shall use the same notation for the estimator and its corresponding estimate.

Proof. As $\frac{SS(bl.adj)}{\sigma 2}$ is distributed as a noncentral chisquare with b-1 degrees of freedom and noncentrality parameter λ^2 (say), we get

$$\frac{E(SS(bl.adj))}{\sigma^2} = (b-1) + \lambda^2$$

So $\lambda^2\sigma^2$ is the noncentrality parameter of the statistic (5.145)

$$F = \frac{(bk-b-\nu+1)}{b-1} \frac{SS(bl.adj)}{q_3} ,$$

for testing H_α; all $\alpha_i = 0$. Since for testing H_α, the power of the test based on F depends on no nuisance parameter except α_i, $i = 1,\dots,b$ and σ^2, we conclude that λ^2 does not depend on μ' and t_j, $j = 1,\dots,\nu$. So to evaluate E(SS(bl.adj)) we can without any loss of generality take $\mu' = 0$, $t_j = 0$ for all j.

Now under the model (5.126) the expression (5.145) is the conditional expected value of $SS(bl.adj)/\sigma^2$ given the a_i and since this depends only on σ^2 and $\alpha = a_i - a_.$, we conclude that the unconditional expected value of $SS(bl.adj)/\sigma^2$ obtained by taking the expected value of the right hand side of (5.145) with respect to the joint distribution of the a_i which are independently normally distributed with mean 0 and the same variance σ^2, will depend on σ^2 and σ_B^2. Under the assumptions $\mu = 0$, $t_j = 0$ for all j, the treatment totals T_j, the adjusted treatment totals Q_j and the block totals B_i all have mean 0. Hence, from (5.144), we get

$$E(SS(bl.adj)) = bk^{-1}Var(B_i) + k\,\lambda^{-1}\,Var(Q_j) - r^{-1}\nu\,Var(T_j)\,. \qquad (5.146)$$

Using (5.117) and (5.126) we get

$$\begin{aligned} Var(B_i) &= \sigma_f^2 = k^2\sigma_B^2 + K\sigma^2, \\ Var(T_j) &= r(\sigma_B^2 + \sigma^2) \\ Var(Q_j) &= Var(r\varepsilon\hat{t}_j) = \frac{r(k-1)\sigma^2}{k} \end{aligned} \qquad (5.147)$$

From (5.146) - (5.147)

$$E(SS(bl.adj)) = (bk-\nu)\sigma_B^2 + (b-1)\sigma^2\,. \qquad \text{Q.E.D.}$$

Hence the unbiased estimate of σ_B^2 is

$$\hat{\sigma}_B^2 = (b-1)(bk-\nu)^{-1}\left(\frac{SS(bl.adj)}{b-1} - \frac{q_3}{(bk-b-\nu+1)}\right) \qquad (5.148)$$

Thus the usual unbiased estimate of σ_f^2 is

$$\hat{\sigma}_f^2 = k(bk-\nu)^{-1}\left[k(b-1)\frac{SS(bl.adj)}{b-1} - \frac{(\nu-k)\,q_3}{(bk-b-\nu+1)}\right] \qquad (5.149)$$

Hence

$$\hat{\psi}^* = \frac{\hat{\omega}\hat{\psi}+\hat{\omega}'\hat{\hat{\psi}}}{\hat{\omega}+\hat{\omega}'}$$

is the required estimate of the contrast ψ using the intrablock and the interblock informations. This is essentially what Yates (1940) did. Graybill and Seshadri (1960) have shown that $\hat{\psi}^*$ is unbiased for ψ and Graybill and Weeks (1959) have shown that ψ^* is based on a set of minimal sufficient statistics. Rao (1947) has derived the maximum likelihood estimates, though Sprott (1956) has shown that this is not true in general for all incomplete block designs.

From (5.150), using (5.110) and (5.133), we get

$$\hat{\psi}^* = \sum_j c_j \hat{t}_j^* \tag{5.151}$$

where

$$\hat{t}_j^* = \frac{\hat{\omega}\,\hat{t}_j+\hat{\omega}'\hat{\hat{t}}_j}{\hat{\omega}+\hat{\omega}'}$$

$$= \frac{\hat{\omega}\,\dfrac{Q_j}{r\varepsilon}+\hat{\omega}'(r-\lambda)^{-1}P_j}{\hat{\omega}+\hat{\omega}'}$$

$$= \frac{k\,Q_j+\hat{\rho}\,P_j}{r(k-1)+\lambda+\hat{\rho}\,(r-\lambda)} \tag{5.152}$$

where

$$P_j = T'_j - rb^{-1}\sum_i B_i\,,$$

$$\hat{\rho} = \frac{rk\,\varepsilon\,\hat{\omega}'}{(r-\lambda)\,\hat{\omega}} = \frac{\hat{\sigma}^2}{k\,\hat{\sigma}_B^2+\hat{\sigma}^2}$$

$$= \frac{v(r-1)\ [(bk-b-v+1)^{-1}\ q_3\ ]}{k\ SS(bl.adj)-(v-k)\,(bk-b-v+1)^{-1}\ q_3} \tag{5.153}$$

where $\hat{\rho}$ is the Yates estimate of

$$\rho = \frac{\sigma^2}{k\sigma_B^2+\sigma^2}$$

From (5.141)

$$\mathrm{Var}(\psi^*) = \sum_j c_j^2 (\omega+\omega')^{-1} \qquad (5.154)$$

Thus

$$\mathrm{Var}(t_j^* - t_{j'}^*) = \frac{2}{\omega+\omega'}$$

$$= \frac{2\sigma^2}{r(k-1)+\lambda+\rho(r-\lambda)} \qquad (5.155)$$

Statistical inferences about the t_j are made by assuming that $\omega=\hat{\omega}$, $\omega'=\hat{\omega}'$ and $\rho=\hat{\rho}$. In other words for all c_j with $\sum_j c_j = 0$, $\hat{\psi}^*$ is normally distributed with mean ψ and with variance given in (5.154) with $\omega=\hat{\omega}$, $\omega'=\hat{\omega}'$.

Example 5.7. An experiment with 13 different varieties of Paddy (rice crop) was conducted at the Agricultural Research Institute, Rajendra Nagar, A.P., India to find the best variety suitable for the tract. A BIB design with $v = b = 13$, $r = k = 4$, $\lambda = 1$ was adopted for the experiment. The layout of the design with yield per plot in kilograms is given in Table 5.11, the number inside brackets being the treatment number.

Table 5.11. Yield of 13 different varieties of paddy per plot in kilograms

Blocks	Yield			
1	24.6 (11)	19.9 (6)	29.0 (9)	25.3 (3)
2	19.8 (4)	33.3 (8)	23.0 (3)	22.7 (12)
3	31.7 (12)	26.6 (13)	19.3 (11)	16.2 (10)
4	27.3 (3)	27.0 (5)	35.6 (8)	17.4 (11)
5	23.4 (7)	30.5 (8)	30.8 (9)	32.4 (10)
6	30.6 (4)	32.4 (5)	27.2 (6)	32.8 (10)
7	34.7 (1)	31.1 (5)	25.7 (9)	30.5 (12)

8	34.4 (3)	32.4 (5)	33.3 (7)	36.9 (13)
9	38.2 (1)	32.9 (2)	37.3 (3)	31.3 (10)
10	28.7 (2)	30.7 (4)	26.9 (9)	35.3 (13)
11	36.6 (1)	31.1 (4)	31.1 (7)	28.4 (11)
12	31.8 (1)	33.7 (6)	27.8 (8)	41.1 (13)
13	30.3 (2)	31.5 (6)	39.3 (7)	26.7 (12)

Table 5.12. Intrablock and combined estimates

Treatments j	$k\ T_j$	T'_j	$k\ Q_j$	Intrablock estimates	Combined estimates $t^*_j = \dfrac{kQ_j + .4517\ P_j}{14.3551}$
1	565.2	523.3	41.9	.805	4.393
2	476.8	496.4	-19.6	-.376	-.738
3	480.0	474.3	5.7	.109	.329
4	448.8	470.6	-21.8	-.419	-1.763
5	491.6	489.3	2.3	.044	.565
6	449.2	484.0	-34.8	-.669	-2.187
7	508.4	509.1	-0.7	-.013	.979
8	508.8	457.6	51.2	.984	2.973
9	449.6	459.5	-9.9	-.190	-1.223
10	450.8	473.6	-22.8	-.438	-1.678
11	358.8	427.1	-68.3	-1.313	-6.311
12	446.4	442.4	4.0	.076	-.793
13	559.6	486.8	72.8	1.400	5.397

Table 5.13. Intrablock analysis of variance

Sources of variation	d.f.	Sum of squares	Mean sum of squares	F
Blocks (ignoring treatments)	12	689.38		
Treatments (eliminating blocks)	12	328.55	27.38	1.374
Error (intrablock)	27	538.22	19.93	
Total	51	1556.15		

The F-value is not significant at the level of significance $\alpha \leq .10$. The estimate of the intrablock variance of $\hat{t}_j - \hat{t}_{j'}$, $j \neq j'$ is

$$\text{Var}(\hat{t}_j - \hat{t}_{j'}) = 12.27$$

$$\text{SS(bl.adj)} = 689.38 - 542.66 + 328.55 = 475.27$$

$$\hat{\rho} = .4517.$$

The combined estimate of treatment effects with recovery of interblock information is

$$\hat{t}^*_j = \frac{k\,Q_j + .4517\,P_j}{14.3551}$$

The estimate of the variance of $\hat{t}^*_j - \hat{t}^*_{j'}$, for $j \neq j'$, is

$$\text{Var}(\hat{t}^*_j - \hat{t}^*_{j'}) = 11.1049\ .$$

5.4.5. Randomization Test in BIB Designs: A BIB design is a plan of a statistical experiment where ν treatments are assigned to experimental plots arranged in b blocks of k plots each such that each treatment occurs exactly once in each of the r blocks and each pair of treatments occurs together in ν blocks. The treatment arrangement in the corresponding plan follows exactly the same rule. For clarity we shall call the blocks of the plan as groups to distinguish them from experimental blocks. The randomization procedure in a BIB design can now be described as follows:

R_1 : the b groups of the plan are distributed at random to the b experimental blocks.

R_2 : Then the k treatments in a block are distributed at random to the k plots of the experimental blocks and that is done separately and independently for each block.

As in the case of the randomized block designs and the latin square designs we shall now investigate the extent to which the normal theory distribution of the statistic $U = Q_2/(Q_2+Q_3)$, where Q_2, Q_3 are random variables corresponding to q_2, q_3 in Table 5.9, can be used to approximate the distribution of U for all possible arrangements under (R_1, R_2) when the hypothesis H_T (all treatments are equal) holds. It may be remarked that this course is only followed when the normality assumptions in (5.100) are in doubt and the hypothesis H_T is tested by using the randomization test. To be exact, one needs to compute the value of U for all possible arrangements of the treatments to the experimental plots. If the value of U, obtained for the layout actually used in the experiment is exceeded in α percent of the cases we reject H_T at the level of significance α. Of course it is a very laborious job and we naturally look for approximations.

Under the model (5.100), the statistic U under H_T has central beta distribution with parameter $(\frac{1}{2}(\nu-1), \frac{1}{2}(bk-b-\nu+1))$. Thus under H_T

$$E(U) = \frac{\nu-1}{b(k-1)} ,$$

$$\begin{aligned} Var(U) &= \frac{2(\nu-1)(bk-b-\nu+1)}{[b(k-1)]^2[b(k-1)+2]} \qquad (5.156)\\ &= \frac{2(\nu-1)(bk-b-\nu+1)}{(b(k-1))^3} \quad \text{if} \quad b(k-1) >> 2,\\ &= \frac{2\nu r(\lambda\nu-k)}{\lambda^2\nu^2 b^2(k-1)} \end{aligned}$$

Mitra (1960) has shown that under the randomization model (R_1, R_2) and under H_T

$$E(U) = \frac{(\nu-1)}{b(k-1)}$$

$$\text{var}(U) = \frac{2r\ (\lambda\nu-k)}{k^2 b^2 (k-1)} \left(1-\frac{\delta}{b-1}\right) \qquad (5.157)$$

where

$$\delta = b \sum_{i=1}^{b} (s_i - \bar{s})^2/s^2$$

$$s_i = \sum_j \left(x_{ij} - \frac{B_i}{k}\right)^2, \qquad s = \sum_{i=1}^{b} s_i, \qquad \bar{s} = \frac{s}{b}.$$

When H_T is true, while the mean of U under (R_1, R_2) agrees perfectly with the mean of U under the model (5.100), the variance of U would also be practically equal if $b(k-1) >> 2$ and the block errors are homogeneous $(\delta=0)$. Here the test based on U (equivalently the analysis of variance F-test) works reasonably well as an approximation to the randomization test. Heterogeneity in block errors reduces to variance of U. Hence if δ is large the approximation by the beta distribution for testing H_T would generally underestimate the significance of departure from H_T and would thus be somewhat conservative.

This situation is similar to that of the randomized block design and in some sense possibly better. When H_T holds the ratio of the variance of U under (R_1, R_2) to that of the corresponding normal theory beta distribution is approximately equal to $1 - \delta/(b-1)$ in a BIB design and to $1 - \delta/(r-1)$ in a randomized block design. In BIB designs $b > r$ and in many cases b is large compared to r.

The types of departure from H_T that can possibly be detected by the normal theory beta test (or equivalently the F-test) have been examined by Rao (1959) through a comparison of the nonnull expected values of the treatment or error mean squares.

5.4.6. Youden Square Designs (position-balanced BIB designs): In the case of a BIB design with $b = \nu$, it is possible to arrange the b blocks and the treatments within the blocks in such a way that they form an incomplete latin square design. In Example 5.6, the following arrangement forms an incomplete latin square design.

Blocks

1	2	3	4	5	6	7
1	2	3	4	5	6	7
2	3	4	5	6	7	1
4	5	6	7	1	2	3
7	1	2	3	4	5	6

Smith and Hartley (1948) have shown that for all incomplete block designs with $b = \nu$ such an incomplete latin square arrangement always exists. These incomplete latin square designs are called Youden square designs after Youden (1940) who first discovered them. In other words a Youden square is a latin square with several rows removed, the remainder having the property that its columns form another BIB design. Cochran and Cox (1950) obtained interesting designs by combining Youden squares with latin squares when the number of treatments is small.

If we treat the position in a block in the BIB design as a third factor then a BIB design is an incomplete 3-way layout, the three factors being blocks at b levels, treatments at ν levels and positions at k levels. It may be desirable to balance out the position and eliminate the position effect (if possible) in such cases rather than randomizing it out where it increases the error variance.

If we impose the restriction on a BIB design with $b = m\nu$ (or equivalently $r = km$) that each treatment appears equally often (m times) in each position, then we obtain a rearrangement into another BIB design which we shall call an m-position balanced BIB design. Obviously an m-position balanced BIB design with $m = 1$ is a Youden square design. Hartley, Shrikhande and Taylor (1953) showed that for any BIB design with $b = m$ there always exists a rearrangement leading to an m-position balanced design.

The linear model assumed for the analysis of a Youden square design is

$$x_{ij\ell} = \mu + \alpha_i + t_j + \beta_\ell + e_{ijk} \qquad (5.158)$$

where $x_{ij\ell}$, $i = 1,\ldots,b$, $j = 1,\ldots,\nu$, $\ell = 1,\ldots,k$ is the observation on the jth treatment in the ith column (block) and ℓth row (position) and μ is the grand mean, α_i is the ith column effect, t_j is the jth treatment effect, β_ℓ is the ℓth position effect and the errors $e_{ij\ell}$ are independently normally distributed with means 0 and common variance σ^2 (and $\sum_i \alpha_i = 0$, $\sum_\ell \beta_\ell = 0$, $\sum_j t_j = 0$). If the interblock information is to be utilized to estimate the treatment effects then it is assumed that the α_i are independently normally distributed with mean 0 and common variance σ_B^2 and are independent of the $e_{ij\ell}$.

The solution has already been obtained (in the BIB design) because the treatments are orthogonal to rows. The estimate is given in (5.152). The intrablock analysis is similar to that of the BIB design except that from the error sum of squares we subtract the sum of squares for rows with k-1 degrees of freedom to get a new error sum of squares with $bk-\nu-b-k+2$ degrees of freedom. The sum of squares for rows is given by $\nu^{-1}\sum_p p_p^2$-C.F., where p_p is the total of the observations in the pth row position.

In the setting of two-way heterogeneity (with additive effects model) with k rows, b columns and ν treatments, we say that a design 'd' is a "generalized Youden square" if it is a balanced block design when rows are considered to be blocks and also when columns are considered to be blocks. A generalized Youden square design is called regular if at least one of k/ν or b/ν is an integer. Kiefer (1958) has proved that a regular generalized Youden square design is A-, D-, E-, and L-optimum. This implies that Youden square designs are A-, D-, E- and L-optimum. If $b = k = \nu$, then a generalized Youden square design is a Latin square design.

5.4.7. Analysis of Covariance for BIB Designs: Cornish (1940) discussed the covariance analysis of BIB design without recovery of interblock information. In the notations of (5.100) the model for the covariance analysis of a BIB design with one concomitant variable z_{ij} can be written as

$$x_{ij} = \mu + \alpha_i + t_j + \gamma z_{ij} + e_{ij} \tag{5.159}$$

where $(i,j) \varepsilon S$. The analysis of covariance for testing H_T: all $t_j = 0$ is given in Table 5.14.

Table 5.14. Covariance analysis of BIB designs without recovery of interblock information

Sources of variation	d.f.	Sum of products			Adjusted S.S.		F
		xx	xz	zz	d.f.	S.S.	
Blocks (ignoring treatment)	$b-1$	B_{xx}	B_{xz}	B_{zz}			
Treatments (eliminating blocks)	$\nu-1$	T_{xx}	T_{xz}	T_{zz}	$\nu-1$	q_4-q_3	$\frac{f-1}{\nu-1}\left(\frac{q_4-q_3}{q_3}\right)$
Error (intrablock)	$f= bk-b-\nu+1$	E_{xx}	E_{xz}	E_{zz}	$f-1$	$q_3=E_{xx}-\frac{E_{xz}^2}{E_{zz}}$	
Treatment (el.blocks) Error$^+$	$f+\nu-1$	W_{xx}	W_{xz}	W_{zz}		$q_4=W_{xx}-\frac{W_{xz}^2}{W_{zz}}$	

where $B_{xz} = k^{-1} \sum_i B_i(x) B_i(z) - b^{-1}k^{-1} G(x) G(z)$,

$$T_{xz} = (r\varepsilon)^{-1} \sum_j Q_j(x) Q_j(z),$$

$$E_{xz} = \sum_{(i,j)\varepsilon S} x_{ij} z_{ij} - k^{-1} \sum_i B_i(x) B_i(z) - (r\varepsilon)^{-1} \sum_j Q_j(x) Q_j(z),$$

$$B_i(x) = \sum_j x_{ij}, \quad B_i(z) = \sum_j z_{ij}, \quad G(x) = \sum_{(i,j)\varepsilon S} x_{ij},$$

$$G(z) = \sum_{(i,j)\varepsilon S} z_{ij},$$

$Q_j(x) = Q_j$, the adjusted jth treatment total with respect to the variables x_{ij}, and similarly for $Q_j(z)$ with respect to the variables z_{ij}, B_{xx}, B_{zz} are the values of B_{xz} when $z = x$ and $x = z$, and similarly for T_{xx}, E_{xx}, etc.

The covariance analysis of Youden square designs can be similarly treated. Covariance analysis in these designs with recovery of interblock information pose several problems. For a detailed treatment we refer to Federer (1963).

5.5. PARTIALLY BALANCED INCOMPLETE BLOCK (PBIB) DESIGNS

We have studied BIB designs in the preceding section. Among equiblock-sized, equireplicated incomplete block designs the BIB designs have several optimum properties. Unfortunately BIB designs do not exist for all combinations of parameters that we might wish to use. Since $\lambda = r(k-1)/(\nu-1)$ is an integer, if $\nu-1$ is a prime number, r has to be a multiple of $\nu-1$ and there is a further restriction that $b = r\nu/k$ has to be an integer.

The main characteristic of a BIB design is that $\mathrm{Var}(\hat{t}_j-\hat{t}_{j'})$ has the same value for all $j \neq j'$. An alternative to BIB designs is to find designs in which this equal variance requirement is modified somewhat. Such designs were introduced by Bose and Nair (1939) and are called partially balanced incomplete block PBIB designs. Their idea has been developed considerably since then mainly by Bose and his associates among others. Bose and Schimamoto (1952) introduced the concept of partially balanced association scheme and rephrased the definition of PBIB designs in terms of this concept.

Definition 5.6. Partially balanced association. Given a set of treatments (symbols) $1,\ldots,\nu$, a relationship satisfying the following three conditions is called a partially balanced association scheme with m associate classes:

(i) Any two treatments in the set are either 1st, 2nd,..., or mth associates and the relation of association is symmetrical, that is, if the treatment α is the ith associate of the treatment β, then β is also the ith associate of the treatment α.

(ii) Each treatment in the set has exactly n_i treatments in the set which are its ith associate and the number n_i does not depend on the treatment chosen.

(iii) If any two treatments α, β are ith associates, then the number of treatments which are both jth associate of α and kth associate of β is p^i_{jk} and is independent of the pair of ith associates chosen. Hence $\mu = p^i_{jk}$. The numbers $\nu, n_1, \dots, n_m$, p^i_{jk} $(i,j,k = 1,\dots,m)$ are called the parameters of the m-associate partially balanced scheme.

Definitions 5.7. A PBIB design with m associate classes is an arrangement of ν treatments into b blocks of size k each, according to an m-associate partially balanced association scheme such that:

a) every treatment occurs at most once in a block,
b) every treatment occurs exactly in r blocks,
c) if two treatments are ith associate, then they occur together exactly in exactly λ_i blocks, the number λ_i being independent of the particular pair of ith associate chosen. It is not necessary that the λ_i should all be different and some of the λ_i may be 0.

The parameters $\nu, b, r, k, \lambda_1, \dots, \lambda_m, n_1, \dots, n_m$ are called by Bose and Nair the parameters of the first kind and the p^i_{jk} are parameters of the second kind.

Obviously

$$bk = \nu r. \qquad (5.160)$$

Since with respect to each treatment, the remaining ν-1 treatments are classified as the 1st, 2nd,...,mth associates and each treatment has n_i ith associates, we have

$$\sum_1^m n_i = \nu - 1 . \qquad (5.161)$$

Consider the r blocks in which a particular treatment ν occurs. In these r blocks we can form r(k-1) pairs of treatments each having α as one of its members. Among these pairs the ith associate of α must occur λ_i times and there are n_i ith associate.

Hence

$$\sum_{1}^{m} n_i \lambda_i = r(k-1). \qquad (5.162)$$

Let the treatments α, β be ith associates. The kth associate of α, $k = 1,\ldots,m$ should contain all the n_j jth associates of $\beta (j \neq i)$. When $j = i$, α itself will be one of the jth associates of β. Hence the kth associates of α, $k = 1,\ldots,m$ should contain all the (n_j-1) jth associates of β. Thus

$$\sum_{k=1}^{m} p^{i}_{jk} = n_j - \delta_{ij} , \qquad (5.163)$$

where $\delta_{ij} = 1$ if $i = j$ and is 0 otherwise.

Let G_i denote the set of ith associates, $i = 1,\ldots,m$ of α. For $i \neq j$, each treatment in G_i has exactly p^{i}_{jk} kth associates in G_j. Similarly each treatment in G_j has exactly p^{j}_{ik} kth associates in G_i. Thus the number of pairs of kth associates that can be obtained by taking one treatment from G_i and another from G_j is on the one hand $n_i p^{i}_{jk}$ and, on the other hand $n_j p^{j}_{ik}$.

Hence

$$n_i p^{i}_{jk} = n_j p^{j}_{ik} = n_k \, p^{k}_{ij} . \qquad (5.164)$$

Among all PBIB designs, the two-associate PBIB designs are most popular in practical applications and most research works, carried out so far, have been concerned with two-associate PBIB designs. We shall be concerned here with two-associate PBIB designs. For construction and other relevant results of PBIB designs we refer to Raghava Rao (1971). Extensive collection of PBIB design with two-associate classes are given by Bose, Clatworthy and Shrikhande (1954). Bose and Schimamoto (1952) classified all the then known PBIB designs with two associate classes in the following five types depending on the association scheme.

(1) Group divisible (or divisible partial designs by Dembowski (1968)), (2) Simple, (3) Triangular, (4) Latin square type and (5) Cyclic.

Definition 5.8. A PBIB design with two associate classes is called group divisible (GD) if $\nu = mn$, where m, n are integers and if these treatments can be divided into m groups of n treatments each such that any two treatments of the same group are first associates and any two treatments of different groups are second associates.

For a GD design the parameters are $\nu = mn$, $n_1 = n-1$, $n_2 = n(m-1)$,

$$P^1 = (p^1_{ij}) = \begin{pmatrix} n-2, & 0 \\ 0, & n(m-1) \end{pmatrix}, \quad P^2 = \begin{pmatrix} 0 & n-1 \\ n-1, & n(m-2) \end{pmatrix}$$

For a GD design $r \geq \lambda_1$, $rk - \lambda_2\nu \geq 0$. The GD designs can be divided into 3 classes (Bose and Connor (1952)), singular designs ($r = \lambda_1$), semiregular designs ($r > \lambda_1$ and $rk = \lambda_2\nu$) and regular designs ($r > \lambda_1$ and $rk > \lambda_2\nu$).

Example 5.8. The following blocks

(15,9,1,13), (5,7,8,1), (10,1,14,2), (15,11,2,13), (6,15,4,7), (12,4,3,1), (12,14, 15, 8), (6,3,14,15), (5,4,2,13), (10,12,13,6), (9,7,10,3), (8,6,2,9), (5,9,11,12), (7,13,14,11), (10,4,8,11) ,

constitute a GD PBIB design with $\nu = b = 15$, $r = k = 4$, $\lambda_2 = 1$, $\lambda_1 = 0$, $P^1 = \begin{pmatrix} 1, & 0 \\ 0, & 12 \end{pmatrix}$, $P^2 = \begin{pmatrix} 0, & 2 \\ 2, & 9 \end{pmatrix}$.

Example 5.9. The following blocks

(1,2,3), (3,4,5), (2,5,6), (1,2,4), (3,4,6), (1,5,6)

form a GD PBIB designs with parameters $\nu = b = 6$, $r = k = 3$, $\lambda_1 = 2$, $\lambda_2 = 1$, $P^1 = \begin{pmatrix} 0, & 0 \\ 0, & 4 \end{pmatrix}$, $P^2 = \begin{pmatrix} 0, & 1 \\ 1, & 2 \end{pmatrix}$. The three groups are (1,2), (3,4), (5,6).

Definition 5.9. A two-associate PBIB design with either $\lambda_1 = 0$, $\lambda_2 \neq 0$, or $\lambda_1 \neq 0$, $\lambda_2 = 0$ is called simple.

A simple PBIB design may be also one of the other types. The GD PBIB design given in Example 5.8 is simple.

Definition 5.10. A two associate PBIB design is called triangular if $\nu = \frac{1}{2} n(n-1)$ and the association scheme is obtained by writing the ν treatments as elements above the principal diagonal of a square array of side n, leaving the principal diagonal blank and repeating the treatments below the diagonal in order to obtain a symmetric array.

For $\nu = 10$, $n = 5$ and we have the following arrangement.

$$\begin{array}{ccccc}
\cdot & 1 & 2 & 3 & 4 \\
1 & \cdot & 5 & 6 & 7 \\
2 & 5 & \cdot & 8 & 9 \\
3 & 6 & 8 & \cdot & 10 \\
4 & 7 & 9 & 10 &
\end{array}$$

For any treatment α, the first associates are exactly those which occur in the same row or the same column as α. Thus the first associates of 1 are (2,3,4,5,6,7) and the second associates of 1 are (8,9,10).

For this association scheme $n_1 = 2(n-2)$, $n_2 = \frac{1}{2}(n-2)(n-3)$

$$P^1 = (p^1_{ij}) = \begin{pmatrix} n-2, & n-3 \\ n-3, & \frac{1}{2}(n-3)(n-4) \end{pmatrix}, \quad P^2 = (p^2_{ij}) = \begin{pmatrix} 4 & , \; 2(n-4) \\ 2(n-4), & \frac{1}{2}(n-4)(n-5) \end{pmatrix}$$

Definition 5.11. A PBIB design with two associate classes is called a latin square type L_i $(i > 2)$ scheme with i constraints if $\nu = s^2$ satisfies the following association scheme: the s^2 ~~treatments~~ treatments in an $s \times s$ square array and i-2 mutually orthogonal ~~latin~~ latin squares are superimposed. Two treatments are first associates if and only if they occur in the same row or the same column of the array or in positions occupied by the same letter in any of the latin squares.

For $s = 4$, taking the array

$$\begin{array}{cccc}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12 \\
13 & 14 & 15 & 16
\end{array}$$

we obtain the following L_2 design with blocks

(1,2,3,4), (5,6,7,8), (9,10,11,12), (13,14,15,16),
(1,5,9,13), (2,6,10,14), (3,7,11,15), (4,8,12,16).

Here $\lambda_1 = 1$, $\lambda_2 = 0$.

Note. L_2 is also a simple lattice design. This we shall discuss in Chapter 6.

Definition 5.12. A non-GD PBIB design with two associate classes is called cyclic if the set of first associates of a treatment α is $(\alpha+d_1,\ldots,\alpha+d_{n_1})$ mod(ν), where the integers $d_1,\ldots,d_{n_1}$ satisfy the following conditions:

1. The d_i are all different and $0 < d_i < \nu$, $i = 1,\ldots,n_1$.
2. Among the $n_1(n_1-1)$ differences, $(d_j-d_{j'})$ (mod(ν)), each of the integers $d_1,\ldots,d_{n_1}$ occurs p^1_{11} times and each of other n_2 positive integers less than ν occurs p^2_{11} times.

All the known cyclic association schemes have parameters

$\nu = 4s+1$, $n_1 = n_2 = 2s$,

$$P^1 = \begin{pmatrix} s-1, & s \\ s, & s \end{pmatrix}, \quad P^2 = \begin{pmatrix} s, & s \\ s, & s-1 \end{pmatrix}.$$

Example 5.10. The following PBIB design with two associate classes with blocks (treatments being numbered as 0,1,2,3,4)

(0,1,2), (1,2,3), (2,3,4), (3,4,0), (4,0,1),

is a cyclic design with $s = 1$, $n_1 = n_2 = 2$, $\lambda_1 = 2$, $\lambda_1 = 1$, $d_1 = 1$, $d_2 = 4$. The difference $d_j - d_{j'}$ are $4-1 = 3$, $1-4 = 2$.

5.5.1. Intrablock Analysis of Two Associate PBIB Designs: We shall consider the intrablock analysis of a PBIB design with parameters ν,b,r,k, λ_1, λ_2, n_1,n_2, p^k_{ij}, $i,j,k = 1,2$, under the normal theory fixed effect model. This is obviously more complicated than that of BIB designs. Let x_{ij} denote the observations on the jth treatment in the ith block provided that the jth treatment occurs in the ith block. Let $N = (n_{ij})$ be the incident matrix of the PBIB design (n_{ij} be the number of times the jth treatment occurs in the ith block). The PBIB design is a binary design implying that $n_{ij} = 1$ or 0. Let S be the set of (i,j) for which $n_{ij} = 1$ and let us assume that the random variables X_{ij} (corresponding to x_{ij}) are independently normally distributed with

$$E(X_{ij}) = \mu + \alpha_i + t_j,$$

$$Var(X_{ij}) = \sigma^2,$$

or

$$x_{ij} = \mu + \alpha_i + t_j + e_{ij} \qquad (5.165)$$

where $(i,j) \varepsilon S$, the e_{ij} are independently normally distributed with mean 0 and variance σ^2, the α_i is the ith block effect satisfying $\sum_i \alpha_i = 0$, the t_j is the jth treatment effect satisfying $\sum_j t_j = 0$. We are interested in testing the null hypothesis H_T: all $t_j = 0$. The hypothesis H_α: all $\alpha_i = 0$, which has no practical significance in most cases, can be similarly treated. In the notation of Section 3.1 the hypothesis H_T is a linear hypothesis with $r = \nu-1$, $s = \nu+b-1$, $N-s = bk-b-\nu+1$. The least squares estimate of μ, α_i, t_j under Ω is obtained by minimizing s^2 in (5.101) and we obtain the set of Equations (5.102) by minimizing s^2 with respect to these parameters. The solution of these equations in the matrix notation is given in (5.122). From (5.122) and (5.124) the estimate of the treatment effects t_j for a general incomplete block design with $k_1 = \dots = k_b = k$, $r_1 = \dots = r_\nu = r$ is $\underline{Q} = C\,\underline{\hat{t}}$, where $C = (rI(\nu) - k^{-1}N'N)$. This can be written as, for $(i,j)\varepsilon S$

$$k\,Q_j = (rk-r)\hat{t}_j - \sum_i \sum_{j'(j'\neq j)} n_{ij}n_{ij'}\hat{t}_{j'}, \quad j = 1,\dots,\nu. \qquad (5.166)$$

Let $S_i(\hat{t}_j)$ be the sum of all estimates of treatments which are ith associates of t_j. Then (5.166) reduces to

$$K\,Q_j = (rk-r)\hat{t}_j - \lambda_1 S_1(\hat{t}_j) - \lambda_2 S_2(\hat{t}_j). \qquad (5.167)$$

Writing $S_i(Q_j)$ for the sum of adjusted treatment totals which are ith associates of t_j, we obtain

$$\begin{aligned} k\,S_1(Q_j) &= -\lambda_1 n_1 \hat{t}_j + S_1(\hat{t}_j)[r(k-1) - \lambda_1 p^1_{11} - \lambda_2 p^1_{22}] \\ &\quad + S_2(\hat{t}_j)\,[-\lambda_1 p^2_{11} \perp \lambda_2 p^2_{12}], \\ k\,S_2(Q_j) &= -\lambda_2 n_2 \hat{t}_j + S_1(\hat{t}_j)\,[-\lambda_1 p^1_{12} - \lambda_2 p^2_{22}] \\ &\quad + S_2(\hat{t}_j)\,[r(k-1) - \lambda_1 p^2_{12} - \lambda_2 p^2_{22}]. \end{aligned} \qquad (5.168)$$

Since

$$\sum_j \hat{t}_j = \hat{t}_j + S_1(\hat{t}_j) + S_2(\hat{t}_j) = 0, \tag{5.169}$$

from (5.167) - (5.169) we obtain

$$k\, Q_j = [rk-r+\lambda_2]\hat{t}_j + (\lambda_2-\lambda_1)S_1(\hat{t}_j),$$

$$\begin{aligned} k\, S_1(Q_j) &= [\lambda_1 p^2_{11} + \lambda_2 p^2_{12} - \lambda_1 n_1]\hat{t}_j \\ &\quad + [r(k-1) + \lambda_1(p^2_{12}-p^1_{12}) + \lambda_2(p^2_{12}-p^1_{22})]S_1(\hat{t}_j) \\ &= [(\lambda_2-\lambda_1)p^2_{12}]\hat{t}_j + [r(k-1)+\lambda_2+(\lambda_2-\lambda_1)(p^1_{11}-p^2_{11})]S_1(\hat{t}_j) \end{aligned} \tag{5.170}$$

as, from (5.163), $p_{11} + p_{12} = n_1-1$, $p^2_{11} + p^2_{12} = n_1$.

From (5.170) we obtain

$$\hat{t}_j = \frac{1}{\Delta}\left(B_{22}k\, Q_j - B_{12}k\, S_1(Q_j)\right), \tag{5.171}$$

where

$$A_{12} = r(k-1) + \lambda_2,\quad B_{12} = \lambda_2 - \lambda_1,\quad A_{22} = (\lambda_2-\lambda_1)p^2_{12},$$

$$B_{22} = r(k-1) + \lambda_2 + (\lambda_2-\lambda_1)(p^1_{11}-p^2_{11}),$$

$$\Delta = A_{12}B_{22} - A_{22}B_{12}. \tag{5.172}$$

Alternatively, if $n_2 < n_1$, it is convenient for computational purpose to eliminate $S_1(\hat{t}_j)$ and obtain

$$\hat{t}_j = \frac{1}{\Delta}\left(B_{21}k\, Q_j - B_{11}k\, S_2(Q_j)\right), \tag{5.173}$$

where

$$A_{11} = r(k-1) + \lambda_1,\quad B_{11} = (\lambda_1-\lambda_2),\quad A_{21} = (\lambda_1-\lambda_2)p^1_{12},$$

$$B_{21} = r(k-1) + \lambda_1 + (\lambda_1-\lambda_2)(p^2_{22}-p^1_{22}),$$

$$\Delta = A_{11}B_{21} - A_{21}B_{11}. \tag{5.174}$$

From (5.106), the error sum of squares for testing H_T in the PBIB design is given by

$$q_3 = [\sum_{(i,j)\in S} x_{ij}^2 - \text{C.F.}] - \sum_j \hat{t}_j Q_j - [k^{-1} \sum_i B_i^2 - \text{C.F.}]$$

$$= [\text{total S.S.}] - [\text{treatment S.S. (eliminating blocks)}]$$

$$- [\text{block S.S. (ignoring treatments)}],$$

where t_j is given by (5.171) or (5.173). The minimum value of s^2 under H_T is equal to

[Total S.S.] - [treatment S.S. (eliminating blocks)]:

The best invariant test of H_T rejects H_T at level α whenever

$$\frac{bk-b-\nu+1}{\nu-1} \left(\frac{\sum \hat{t}_j Q_j}{q_3} \right) \geq F_{1-\alpha,\nu-1,bk-\nu-b+1}. \tag{5.175}$$

The variance of the estimator $\hat{t}_j - \hat{t}_{j'}$ of the elementary treatment contrast $t_j - t_{j'}$ (based on intrablock information) is given by (from (5.171))

$$\operatorname{Var}(\hat{t}_j - \hat{t}_{j'}) = \begin{cases} 2\,\Delta^{-1} k (B_{22} + B_{12}) \sigma^2, & \text{if } j \text{ and } j' \text{ are first associates,} \\ 2\,\Delta^{-1} k\, B_{22}\, \sigma^2, & \text{if } j \text{ and } j' \text{ are second associates.} \end{cases} \tag{5.176}$$

The intrablock analysis of variance of a 2-associate PBIB design is given in Table 5.15.

Table 5.15. Intrablock analysis of variance of a two-associate PBIB design.

Sources of variation	d.f.	Sum of squares	F
Blocks (ignoring treatments)	$b-1$	$k^{-1} \sum_i B_i^2 - \text{C.F.}$	
Treatments (eliminating blocks)	$\nu-1$	$q_2 = \sum_j \hat{t}_j Q_j$	$\frac{bk-b-\nu+1}{\nu-1}\ \frac{q_2}{q_3}$
Error	$bk-b-\nu+1$	q_3	
Total	$bk-1$	$\sum_{(i,j)\in S} x_{ij}^2 - \text{C.F.}$	

Note C.F. $= b^{-1}k^{-1}[\sum_S x_{ij}]^2$.

If the F-value is significant, using (5.176) we can find which of the elementary contrast in the treatment effects are significantly different from 0.

5.5.2. Missing Observations in PBIB Design: One can treat the case of one missing observation in the two-associate class PBIB design in the same way as the BIB design in 5.4.2. Giri (1957) discussed the analysis of several missing observations when the missing observations follow a particular pattern. In this context he has defined a new class of design which he called "reinforced PBIB designs".

5.5.3. The Efficiency of the Two-associate PBIB Design: From (5.176) the average variance of the elementary treatment contrasts, using intrablock information only, is

$$2(\nu-1)^{-1} [n_1 k\Delta^{-1}(B_{22}+B_{12}) + n_2 k\Delta^{-1}B_{22}] \sigma^2 \tag{5.177}$$

If we have conducted the experiment in the randomized block design with the same number of treatments ν, the same number of replication r for each treatment and the same σ^2, then the variance of the unbiased estimator of any elementary treatment contrast $t_j - t_{j'}$ is $2\sigma^2/r$. So the efficiency of the PBIB design (2-associate class) relative to the randomized block design is

$$\frac{(\nu-1)\Delta}{rn_1k(B_{22}+B_{12})+rn_2kB_{22}} = \frac{(\nu-1)\Delta}{rk[(\nu-1)B_{22}+n_1B_{12}]} . \tag{5.178}$$

5.5.4. Recovery of Intrablock Information: The recovery of interblock information in general incomplete block designs has been discussed in 5.4.4. As discussed there, the combined estimates suggested by Yates and the maximum likelihood estimates of Rao are identical in BIB designs. But, this is not in general true for PBIB designs. Sprott (1956) has shown that these two methods give the same combined estimate for treatment effects only for group divisible two-associate PBIB designs.

We shall assume the model (5.126) for the recovery of interblock information in PBIB designs. The intrablock estimates of treatment effects, obtained by minimizing s^2 in (5.130), is given in (5.131). The combined (intrablock and interblock) estimate of t_j is obtained by minimizing

$$W_1 \sum_{(i,j)\varepsilon S} (x_{ij} - \mu - \alpha_i - t_j)^2 + W_1' k^{-1} \sum_i \Big(B_i - k\mu' - \sum_j n_{ij} t_j\Big)^2$$

with respect to parameters μ, μ', α_i, t_j, where

$$W_1 = \frac{1}{\sigma^2}, \quad W_1' = \frac{1}{\sigma_B^2 + k\sigma^2} \qquad (5.179,$$

are assumed to be known, so that if we combine (5.105) and (5.131) with appropriate weights, the combined estimates $\hat{t}_j^*$ of the treatment effects t_j are given by

$$r\,\Big[W_1\Big(\frac{k-1}{k}\Big) + \frac{W'_1}{k}\Big]\hat{t}^*_j - \Big(\frac{W_1-W'_1}{k}\Big) \sum_{j'(\neq j)} \lambda_{jj'}\, \hat{t}^*_{j'}$$

$$= W_1Q_j + \frac{W_1'}{k} P_j, \quad j = 1,\dots,\nu, \qquad (5.180)$$

where $P_j = T'_j - rb^{-1}\sum_i B_i$. Writing

$$\alpha = r\Big[W_1\Big(\frac{k-1}{k}\Big) + \frac{W'_1}{k}\Big], \quad \beta = \frac{W_1 - W'_1}{k}$$

$$V_j = W_1Q_j + \frac{W'_1}{k}P_j \qquad (5.181)$$

and comparing (5.180) with (5.166) we can write, as in (5.167)

$$\alpha\,\hat{t}_j^* - \beta\lambda_1 S_1(\hat{t}^*_j) - \beta\lambda_2 S_2(\hat{t}^*_j) = V_j, \quad j = 1,\dots,\nu . \qquad (5.182)$$

Following (5.168), we obtain from (5.182)

$$\begin{aligned} &-\beta\lambda_1 n_1 \hat{t}^*_j + (\alpha - \beta\lambda_1 p^1_{11} - \beta\lambda_2 p^1_{12}) S_1(\hat{t}^*_j) \\ &+ (-\beta\lambda_1 p^2_{11} - \beta\lambda_2 p^2_{12})\; S_2(\hat{t}^*_j) = S_1(V_j), \\ &-\beta\lambda_2 n_2 \hat{t}^*_j + (-\beta\lambda_1 p^1_{21} - \beta\lambda_2 p^1_{22}) S_1(\hat{t}^*_j) \\ &+ (\alpha - \beta\lambda_1 p^2_{21} - \beta\lambda_2 p^2_{22}) = S_2(V_j). \end{aligned} \qquad (5.183)$$

Since $\hat{t}^*_j + S_1(\hat{t}^*_j) + S_2(\hat{t}^*_j) = 0$, from (5.182)-(5.183), we obtain

$$A'_{12}\,\hat{t}^*_j + B'_{12}S_1(\hat{t}^*_j) = kV_j,$$

$$A'_{22}\,\hat{t}^*_j + B'_{22}S_1(\hat{t}^*_j) = kS_1(V_j)\ , \qquad (5.184)$$

where

$$A'_{12} = k(\alpha+\beta\lambda_2) = r[W_1(k-1)+W'_1] + (W_1-W'_1)\lambda_2,$$

$$B'_{12} = k\beta(\lambda_2-\lambda_1) = (\lambda_2-\lambda_1)\ (W_1-W'_1)\ ,$$

$$A'_{22} = k\beta\lambda_2 p^2_{12} - k\beta\lambda_1 p^2_{12} = (W_1-W'_1)(\lambda_2-\lambda_1)p^2_{12}, \qquad (5.185)$$

$$B'_{22} = k - k\beta\lambda_1(p^1_{11}-p^2_{11}) - k\beta\lambda_2(p^1_{12}-p^2_{12})$$

$$= r[W_1(k-1)+W'_1] + (W_1-W'_1)\ [\lambda_2+(\lambda_2-\lambda_1)(p^1_{11}-p^2_{11})]$$

as $p^1_{11} + p^1_{12} = n_1-1$, $p^2_{11} + p^2_{12} = n_1$. From (5.184), we obtain

$$\hat{t}^*_j = \frac{1}{\Delta}\,[B'_{22}kV_j - B'_{12}\ kS_1(V_j)]\ , \qquad j = 1,\ldots,\nu\ , \qquad (5.186)$$

where $\Delta' = A'_{12}\,B'_{22} - A'_{22}\,B'_{21}$. If $n_2 < n_1$, computationally it is convenient to solve the equations (obtained from (5.183) by eliminating $S_1(\hat{t}^*_j)$).

$$A'_{11}\,\hat{t}^*_j + B'_{11}\,S_2(\hat{t}^*_j) = k\ V_j,$$

$$A'_{21}\,\hat{t}^*_j + B'_{21}S_2(\hat{t}^*_j) = kS_2(V_j), \qquad (5.187)$$

and obtain

$$\hat{t}^*_j = \frac{1}{\Delta'}\,[B'_{21}kV_j - B'_{11}\ kS_2(V_j)], \quad j = 1,\ldots,\nu\ , \qquad (5.188)$$

where

$$A'_{11} = r[W_1(k-1)+W'_1] + (W_1-W'_1)\lambda_1, \quad B'_{11} = (W_1-W'_1)(\lambda_1-\lambda_2),$$

$$A'_{21} = (W_1-W'_1)(\lambda_1-\lambda_2)p'_{12}\ ,$$

$$B'_{21} = r[W_1(k-1)+W'_1]+(W_1-W'_1)[\lambda_1+(\lambda_1-\lambda_2)(p^2_{22}-p^1_{22})],$$

$$\Delta' = A'_{11}B'_{21} - A'_{21}B'_{11} = A'_{12}B'_{22} - A'_{22}B'_{21}.$$

It may be verified, from (5.186), that

$$\mathrm{Var}(\hat{t}^*_j-\hat{t}^*_{j'}) = \begin{cases} \frac{2k}{\Delta'}(B'_{12}+B'_{22})\sigma^2, & \text{if } (j,j') \text{ are first associates,}\\ \frac{2k}{\Delta'}B'_{22}\,\sigma^2 & \text{if } (j,j') \text{ are second associates.}\end{cases} \tag{5.189}$$

As in the case of BIB designs W_1, W'_1 are unknown. Their estimates $\hat{W}_1, \hat{W}'_1$ are obtained from (5.142) and (5.143). The estimate $\hat{\rho}$ of $\rho = W'_1/W_1$ is given in (5.153). Statistical inferences about the t'_j are made by assuming that $W_1 = \hat{W}_1$, $W'_1 = \hat{W}'_1$ and $\rho=\hat{\rho}$.

Example 5.11. An experiment with 15 different varieties of wheat was conducted at Indian Agricultural Research Institute, New Delhi, India, to find the best variety suitable for the region. A group divisible PBIB design with $v = b = 15$, $r = k = 4$, $\lambda_1 = 1$, $\lambda_2 = 0$, $n_1 = 12$, $n_2 = 2$,

$$P^1 = (p^1_{ij}) = \begin{pmatrix} 9, & 2 \\ 2, & 0 \end{pmatrix}, \quad P^2 = (p^2_{ij}) = \begin{pmatrix} 12, & 0 \\ 0, & 1 \end{pmatrix}$$

was adopted for the experimental layout. The layout of the design with yield per plot in kilograms is given in Table 5.16, the number inside brackets being the treatment number.

Table 5.16. Yield of 15 different varieties of wheat

Blocks	Yields with treatments			
1	2.4 (15)	2.5 (9)	2.6 (1)	2.0 (13)
2	2.7 (5)	2.8 (7)	2.4 (8)	2.7 (1)
3	2.6 (10)	2.8 (1)	2.4 (14)	2.2 (2)
4	3.4 (15)	3.1 (11)	2.1 (2)	2.3 (3)
5	4.1 (6)	3.3 (15)	3.3 (4)	2.9 (7)
6	3.4 (12)	3.2 (4)	2.8 (3)	3.0 (1)

7	3.2 (12)	2.5 (14)	2.4 (15)	2.6 (8)
8	2.3 (6)	2.3 (3)	2.4 (14)	2.7 (5)
9	2.8 (5)	2.8 (4)	2.6 (2)	2.5 (13)
10	2.5 (10)	2.7 (12)	2.8 (13)	2.6 (6)
11	2.6 (9)	2.6 (7)	2.3 (10)	2.4 (3)
12	2.7 (8)	2.7 (6)	2.5 (2)	2.6 (9)
13	3.0 (5)	3.6 (9)	3.2 (11)	3.2 (12)
14	3.0 (7)	2.8 (13)	2.4 (14)	2.5 (11)
15	2.4 (10)	2.5 (4)	3.2 (8)	3.1 (11)

From (5.173), for intrablock estimates

$A_{11} = 13,\ B_{11} = 1,\ A_{21} = 2,\ B_{21} = 14,\ \Delta = 180.$

$$\hat{t}_j = \frac{k\,Q_j - k\,S_2(Q_j)}{180}, \quad j = 1, \ldots, \nu .$$

The combined estimates $\hat{t}^*_j$, from (5.188), can be written as

$$\hat{t}^*_j = \frac{(a+b)R_j - b\,S_2(R_j)}{a^2+ab+2b^2}$$

where

$$R_j = k\,Q_j + \rho P_j, \quad a = rk - (1-\rho)(r-1), \quad b = 1-\rho,$$

$$\rho = \frac{W'_1}{W_1},$$

and $S_2(R_j)$ is the sum of the R_j of all treatments which are second associates of t_j.

Calculation for combined estimates. Treatment sum of squares (ignoring blocks)

$$= r^{-1} \sum_j T_j^2 - \text{C.F.} = 3.1383 .$$

Table 5.17. Intrablock estimates

Treatment j	T_j	T'_j	kQ_j	2nd associates	$kS_2(Q_j)$	$\hat{t}_j$
1	11.1	42.5	1.9	6, 11	4.2	0.1244
2	9.4	42.1	-4.5	7, 12	3.7	-0.3700
3	9.8	42.9	-3.7	8, 13	-0.5	-0.2850
4	11.8	47.9	-0.7	9, 14	0	-0.0544
5	11.2	44.0	0.8	10, 15	-1.2	0.0688
6	11.7	44.4	2.4	11, 1	3.7	0.1661
7	11.3	44.8	0.4	12, 2	-1.2	0.0377
8	10.9	43.0	0.6	13, 3	-4.8	0.0733
9	11.3	42.9	2.3	14, 4	-3.0	0.1955
10	9.8	41.7	-2.5	15, 5	2.1	-0.2061
11	11.9	45.8	1.8	1, 6	4.3	0.1161
12	12.5	46.7	3.3	2, 7	-4.1	0.2794
13	10.1	41.5	-1.1	3, 8	-3.1	-0.0683
14	9.7	41.1	-2.3	4, 9	1.6	-0.1877
15	11.5	44.7	1.3	5, 10	-1.7	0.1105

Table 5.18. Intrablock analysis of variance

Sources of variations	d.f	S.S.	Mean S.S.	F
Blocks (ignoring treatments)	14	4.9233	0.3516	
Treatments (eliminating blocks)	14	1.5632	0.1150	1.328
Error (intrablock)	31	2.6878	0.0867	
Total	59	9.1743		

Block sum of squares (eliminating treatments)

= [Block sum of squares (ignoring treatments)] - [treatment sum of squares (ignoring blocks)] + [treatment sum of squares (eliminating blocks)]

= 4.9233 - 3.1383 + 1.5632

= 3.3482 .

$$\text{Block mean square (adjusted)} = \frac{3.3482}{14} = 0.2420 .$$

$$\hat{\rho} = \frac{\hat{W}_1'}{\hat{W}_1} = \frac{45(0.0867)}{56(0.2420)-11(0.0867)} = 0.3099 ,$$

$$a = 13.9297, \quad b = 0.6901, \quad a^2 + ab + 2b^2 = 204.6019.$$

$$\hat{t}_j^* = \frac{14.6198\, R_j - 0.6901\, S_2(R_j)}{204.6019} , \quad j = 1, \ldots, v.$$

$$\mathrm{Var}(\hat{t}_j^* - \hat{t}_{j'}^*) = \begin{cases} 0.0519, & \text{if } (j,j') \text{ are second associates,} \\ 0.0495, & \text{if } (j,j') \text{ are first associates.} \end{cases}$$

Table 5.19. Estimates of $\hat{t}_j^*$

Treatments j	$\hat{t}_j^*$
1	0.0877
2	-0.3746
3	-0.2782
4	0.0457
5	0.0681
6	0.1728
7	0.0567
8	0.0459
9	0.1543
10	-0.2322
11	0.1603
12	0.3158
13	-0.1161
14	-0.2317
15	0.1217

5.5.5. Randomization Test in PBIB Designs: The reader is referred to Section 5.4.5 for various details and for terminology used in this section. We shall investigate here the extent to which the normal theory beta distribution can be used to approximate the distribution of U under (R_1,R_2) and under H_T: all $t_j = 0$ for two associate PBIB designs. This is done by computing the first two moments of U under (R_1,R_2) and comparing them with the corresponding moments of a central beta random variable with parameters $(\frac{1}{2}(\nu-1), \frac{1}{2}(bk-b-\nu+1))$. For two associate PBIB designs, the first two moments of U under (R_1,R_2) were computed by Rao (1963) for the case $\lambda_1 = 0$, $\lambda_2 = 1$, by Giri (1965) for the case $\lambda_1 = 0$, by Cleroux (1969) for the general case. In all cases $E(U) = b^{-1}(k-1)^{-1}(\nu-1)$, which agrees with E(U) under the normal theory model (5.165). The computation of Var(U) under (R_1,R_2) is much more complicated and no direct analytical comparison can be made with Var(U) under the model (5.165). Cleroux (1969), with aid of computer, computed the ratio R of the variance of U under (R_1,R_2) to the variance of U under the model (5.165) for nearly two hundred designs and made the following observations:

(1) For group divisible design these two variances are approximately equal, that is, $90 \leq R \leq 1.10$.

(2) For simple designs these two variances are approximately equal in some specific cases.

(3) In all other cases they are unequal.

However, in cases where these two variances are unequal one can approximate the distribution of U under (R_1,R_2) and H_T by a central beta random variable whose degrees of freedom are such that the mean and the variance of these two distributions are equal respectively. (We refer to Cleroux (1969) for the required expressions for the degrees of freedom in terms of the first two columns of U under R_1,R_2.

5.5.6. Row-balance (Position-balance) in PBIB Designs: The entire class of PBIB designs can be divided into 3 broad classes depending on (i) $b = \nu$, (ii) $b = m\nu$, (iii) $b \neq m\nu$ or $r = mk+t$.

where $m > 1$ and τ are positive integers. Bose and Kishen (1939) developed the row-balancing in PBIB designs with $b = \nu$ and discussed their analysis without recovery of interblock information. Shrikhande (1951) extended this idea to the case $b = m\nu$ as well as $b \neq m\nu$ and found two designs with $b \neq m\nu$ for which row-balancing is possible. Giri (1957) obtained a class of designs with $b \neq m\nu$ for which row-balancing is possible. As $r \neq mk$, arrangement of treatments within blocks so that every treatment is replicated a constant number of times in each row is impossible. The alternative is that some partially balanced arrangement of the treatments within rows may be possible. It may be possible to have such arrangements for many PBIB designs but an analysis after eliminating row effects becomes straightforward if the following conditions are satisfied:

(i) each treatment is replicated at least m times and atmost (m+1) times in each row;

(ii) the (m+1)th replicate of the treatments (named as odd treatments) can be so arranged among the rows that the rows form a PBIB design.

Example 5.12. Consider a PBIB design with $\nu=12$, $b=16$, $k=3$, $r=4$, $\lambda_1=1$, $\lambda_2=0$, $n_1=8$, $n_2=3$,

$$P^1 = \begin{pmatrix} 4, & 3 \\ 3, & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 8, & 0 \\ 0, & 2 \end{pmatrix},$$

having the following arrangement .

	Blocks															
Rows	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	2	5	7	10	4	1	6	3	11	8	12	9	10	7	4	1
	1	4	8	11	12	9	10	7	6	3	5	2	5	2	11	8
	3	6	9	12	8	11	2	5	7	10	1	4	9	12	3	6

The arrangement of the odd treatments in the rows are

Rows	Odd treatments in the rows			
1	1	4	7	10
2	2	5	8	11
3	3	6	9	12

which is the PBIB design with parameters, $v = 12$, $b = 3$, $r = 1$, $k = 4$, $n_1 = 8$, $n_2 = 3$, $\lambda_1 = 0$, $\lambda_2 = 1$, P^1, P^2.

Giri (1957) has proved that the following class of designs given by Bose and Nair (1939) with parameters $v = p(p-1)$, $b = p^2$, $r = p$, $k = p-1$, $n_1 = p(p-2)$, $n_2 = p-1$, $\lambda_1 = 1$, $\lambda_2 = 0$,

$$P^1 = \begin{pmatrix} p(p-3), & b-1 \\ p-1 & , \quad 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} p(p-2) & , \quad 0 \\ 0 & , \quad p-2 \end{pmatrix}$$

where p is a prime or a power of a prime number can be arranged satisfying the above two conditions. Example 5.12 is a special case of it when p = 4. The analysis with and without recovery of interblock information is quite straightforward and goes in the same way as the row-balanced BIB design in 5.4.6. The linear model assumed for the analysis is given in (5.158). The reader is referred to Giri (1957) for details.

EXERCISE 5

1. An agricultural field experiment was conducted on cotton at the Dry Farming Research Station at Dhanduka, India, with the object ~~object~~ of finding out a suitable level of farm yard manure for a maximum ~~maximum~~ yield of cotton. The treatments are, (1) no farm yard manure, (2) 15 tonnes of farm yard manure per hectare, (3) 30 tonnes of farm yard manure per hectare, (4) 75 tonnes of farm yard manure per hectare, (5) 150 tonnes of farm yard manure per hectare and (6) 300 tonnes of farm yard manure per hectare. The layout (completely randomized design) with yield of cotton per plot in kilograms is given below.

The number inside the brackets indicates the treatment number. Analyse the experimental data to test the hypothesis that the treatment effects are equal.

5.40 (4)	8.90 (6)	5.85 (2)	3.10 (1)	5.05 (3)	2.25 (2)
8.40 (4)	4.50 (1)	3.90 (6)	6.40 (5)	3.60 (2)	6.70 (3)
3.20 (3)	3.40 (6)	4.45 (6)	8.05 (6)	6.70 (4)	5.40 (5)
5.50 (4)	5.80 (5)	3.70 (1)	4.50 (3)	5.65 (1)	4.30 (2)

2. Suppose that in a randomized block design with ν treatments and r replications for each treatment the treatment i in the block j is interchanged with the treatment j in the block i. Give the analysis of variance of the resulting design.

3. In a randomized block design suppose that treatment 1 was used twice by mistake in the first two blocks. How do you propose to analyse the resulting experiment?

4. Suppose that in a $\nu \times \nu$ Latin square design two treatments in the same row (or column) got interchanged by error. Obtain the unbiased estimator (if any) of different elementary treatment contrasts and their variances.

5. An experiment in a Latin square layout was conducted at the Agricultural Research Station, Marutern, Andhra Pradesh, India, to find out the different bulky organic manures, each supplying 40 kilograms of Nitrogen per hectare, on the yield of Paddy. The treatments are,

(1) farm yard manure, (2) Glyricidia, (3) farm waste compost,
(4) night soil compost, (5) sannhemp compost.

The actual experimental layout with yield of paddy per plot in kilograms is given below.

12.5 (3)	16.3 (5)	14.0 (1)	8.5 (4)	10.2 (3)
14.5 (5)	13.5 (1)	7.9 (4)	9.2 (2)	11.5 (3)
13.5 (1)	10.5 (4)	8.5 (2)	13.6 (3)	18.9 (5)
11.0 (4)	13.2 (2)	15.2 (3)	20.3 (5)	15.6 (1)
14.5 (1)	10.5 (3)	12.0 (5)	10.9 (1)	8.8 (4)

Analyse the data to find the best treatment (if any).

6. Show that for a symmetrical BIB design with ν even, $(r-\nu)$ must be a perfect square.

7. Let N be the incident matrix of a BIB design with parameters ν, b, r, k, λ.

(a) Show that $|NN'| = 0$ if $b \neq \nu$.

(b) Show that the characteristic roots of NN' are rk and $r-\lambda$ of multiplicaties 1 and $\nu-1$ respectively.

(c) Show that if $b = \nu$

$$|NN'| = r^2(r-\lambda)^{\nu-1} .$$

8. Prove that for a resolvable BIB design rank $N \leq b-(r-1)$.

9. Given a set of ν treatments, consider a BIB design with parameters ν, $b = \frac{1}{2}\nu(\nu-1)$, $k = 2$, $r = \nu-1$, $\lambda = 1$ (blocks are formed by taking all possible different pairs of treatments). Taking blocks for treatments and treatments for blocks in this design obtain a new design in which any two blocks have a single treatment as a common link. Show that the new design is a PBIB design. Obtain the parameters of this design.

10. The followings are the plan and yield in lbs of 13 varieties of wheat (numbered 1 to 13) in an experimental trial conducted in the Agricultural Research Institute, Pusa, India. The figures inside the brackets indicate the variety number.

Blocks	Yield of wheat by varieties			
1	50.1 (3)	39.5 (6)	57.3 (9)	48.5 (11)
2	45.3 (3)	38.9 (4)	66.7 (8)	45.3 (12)
3	32.0 (10)	38.2 (11)	63.5 (12)	53.9 (13)
4	54.1 (2)	54.5 (5)	70.7 (8)	34.3 (11)
5	46.5 (7)	60.9 (8)	61.3 (9)	64.3 (10)

6	60.7 (4)	65.0 (5)	54.2 (6)	65.1 (10)
7	68.5 (1)	62.1 (5)	51.7 (9)	61.3 (12)
8	68.5 (3)	65.1 (5)	66.1 (7)	72.9 (13)
9	76.7 (1)	65.3 (2)	75.1 (3)	62.6 (10)
10	57.1 (2)	60.9 (4)	53.5 (9)	70.3 (13)
11	72.9 (1)	62.3 (4)	62.7 (7)	57.1 (11)
12	63.9 (1)	67.3 (6)	55.3 (8)	84.3 (13)
13	60.7 (2)	63.1 (6)	79.1 (7)	53.5 (12)

Identify the layout actually used in this experiment. Analyse the data to find the best variety if any.

11. With a view to selecting the varieties of wheat giving higher yields an experiment was conducted with 9 varieties of wheat in Haryana Agricultural University using a symmetrical PBIB design. The layout of the design with yield per plot in kilogram is given below. The figures inside the brackets indicate variety number.

Blocks	Yield in kilograms			Blocks	Yield in kilograms		
1	59 (3)	56 (8)	53 (4)	6	52 (3)	53 (9)	48 (5)
2	35 (2)	33 (7)	40 (4)	7	54 (1)	59 (8)	62 (6)
3	48 (1)	42 (7)	43 (5)	8	45 (2)	46 (9)	47 (6)
4	46 (7)	56 (8)	50 (9)	9	31 (1)	28 (2)	25 (3)
5	61 (4)	61 (5)	54 (6)				

Find the parameters of the design. Analyse the data.

12. Consider a BIB design with parameters ν, b, r, k, λ. Suppose a control treatment is added to each block so that the block size is now increased to $k+1$. How do you plan to analyse such a design?

13. Source: Biometrics Vol.11, 1955, page 112, edited by G.W.Snedecor.

An experiment was conducted in a Latin square layout to study the responses of five pairs of monkeys to a certain kind of stimulus under five different conditions: A,B,C,D,E during five periods consisting of successive weeks. The figures inside the brackets represent the total number of responses.

Total number of responses

Animal	Period				
	1	2	3	4	5
1	B(194)	D(369)	C(344)	A(380)	E(693)
2	D(202)	B(142)	A(200)	E(356)	C(473)
3	C(335)	A(301)	E(439)	B(338)	D(528)
4	E(515)	C(590)	B(552)	D(677)	A(546)
5	A(184)	E(421)	D(355)	C(284)	B(366)

(a) Analyse the data under normal theory model.

(b) Analyse the data assuming that the value in period 2 with animal 2, and the value in period 4 with animal 4 are missing.

14. Source: Introduction to Engineering Statistics by I. Guttman, the late S.S.Wilks and J.S.Hunter, John Wiley(1971).

The following data were recorded in a radioactive counting rate experiment of 4 specimens: S_1,S_2,S_3,S_4 of radium. These specimens were subjected to a counter with 4 shielding methods: M_1,M_2,M_3,M_4 in 4 different orders: R_1,R_2,R_3,R_4 so that the entire experiment layout is a Latin square design.

Shielding Method

Order	M_1	M_2	M_3	M_4
R_1	S_1(26.46)	S_3(29.61)	S_2(27.82)	S_4(29.15)
R_2	S_2(27.58)	S_4(29.52)	S_1(26.48)	S_3(29.13)
R_3	S_3(29.54)	S_1(27.00)	S_4(29.31)	S_2(27.90)
R_4	S_4(29.51)	S_2(28.03)	S_3(29.53)	S_1(26.51)

Carry out the analysis of variance. Use S- and T-Methods with $\alpha = .05$ to compare pairwise the different shielding methods.

15. Source: Introduction to Engineering Statistics, by I.Guttman, the late S.S. Wilks and J.S. Hunter, John Wiley, (1971).

Microminature integrated circuits are manufactured, in part, by decompositing thin films of dialectric material in predesigned patterns. Prior to mounting the film, the subtrate has to be prepared. In a study to determine the best operating conditions for preparation 4 factors: A,B,C,D, each at 2 levels are studied. The experiment was blocked into 2 blocks of 8 runs each and each block corresponding to a day. The design and the deposited SIO film (y) are given below.

A = Subtrate temperature, levels, $A_o = 250^oC$, $A_1 = 300^oC$,

B = Vacuum in chamber 1, mm of Hg,
levels, $B_o = 1 \times 10^{-5}$,
$B = 1.5 \times 10^{-5}$.

C = Vacuum in chamber 2, mm of Hg,
levels, $C_o = 1 \times 10^{-5}$, $C_1 = 1.5 \times 10^{-5}$.

D = Pattern types, levels, D_o = type 1, D_1 = type 2.

Block 1		Block 2	
Treatment	y	Treatment	y
$A_oB_oC_oD_o$	3.43	$A_oB_oC_oD_1$	3.62
$A_1B_oC_oD_1$	4.04	$A_1B_oC_oD_o$	4.17
$A_oB_1C_oD_1$	3.57	$A_oB_1C_oD_o$	3.48
$A_1B_1C_oD_o$	3.86	$A_1B_1C_oD_1$	4.18
$A_oB_oC_1D_1$	4.09	$A_oB_oC_1D_o$	3.27
$A_1B_oC_1D_o$	3.27	$A_1B_oC_1D_1$	4.35
$A_oB_1C_1D_o$	3.15	$A_oB_1C_1D_1$	4.09
$A_1B_1C_1D_1$	4.20	$A_1B_1C_1D_o$	3.52

Analyse the data and test the hypothesis that there is no difference of block means.

16. Source: An application of a balanced incomplete block design, P.W.M. John, 1961, Technometrics.

The data were collected from a dishwashing experiment where the treatments were detergents and the response observed was the number of plates washed under standard test conditions before the foam disappeared. In the experiment the detergent solutions are made up and plates soiled with standard soil were washed one at a time until they are clean. The test procedures calls for 3 basins to be used for 3 treatments to be tested simultaneously and the operators wash at a common speed during a test. A block is thus a set of 3 operators testing 3 detergent simultaneously. The treatments are denoted by A,...,I and the responses are given within brackets.

Blocks				Blocks			
1	A(19)	B(17)	C(11)	7	A(20)	E(26)	I(31)
2	D(6)	E(26)	F(23)	8	B(16)	F(23)	G(21)
3	G(21)	H(19)	I(28)	9	C(13)	D(7)	H(20)
4	A(20)	D(7)	G(20)	10	A(20)	F(24)	H(19)
5	B(17)	E(26)	H(19)	11	B(17)	D(6)	I(29)
6	C(15)	F(23)	I(31)	12	C(14)	E(24)	G(21)

Analyse the data and interpret the results.

REFERENCES

Allan, F.E. and Wishart, J. (1936). A method of estimating yield of a missing plot in field experimental work, J. Agri. Sci., 20, 399-406.

Bose, R.C. (1938). On the application of the properties of Galois field to the construction of Hyper- Graeco Latin Squares, Sankhya, 3, 323-338.

Bose, R.C.((1939). On the construction of balanced incomplete block design, Ann. Eugen. 9, 353-399.

Bose, R.C. (1942a). On some new series of balanced incomplete block designs, Bull. Calcutta Math. Soc., 34, 17-31.

Bose, R.C. (1942b). A note on the resolvability of incomplete block designs, Sankhya, 6, 105-110.

Bose, R.C. (1942c). A note on two series of balanced incomplete block designs, Bull. Calcutta Math. Soc., 34, 129-130.

Bose, R.C. (1947). On a resolvable series of balanced incomplete block designs. Annal. Math. Statist., 20, 619-620.

Bose, R.C. (1959). On the application of finite projective geometry for deriving a certain series of balanced Kirkman arrangements, Calcutta Math. Soc. Golden Jubilee Vol., 341-354.

Bose, R.C., Chatworthy, W.H. and Shrikhande, S.S. (1954). Tables of partially balanced incomplete block designs with two associate classes, North Carolina Agricultural experimental bulletin No. 167.

Bose, R.C. and Connor, W.S. (1952). Combatorial properties of group divisible incomplete block designs. Ann. Math. Statist., pp.347-383.

Bose, R.C. and Kishen, K. (1939). On partially balanced Youden's squares, Sci. Culture, 5, 136-137.

Bose, R.C. and Nair, K.R. (1939), Partially balanced incomplete block designs, Sankhya, 4, 337-372.

Bose,R.C. and Schimamoto, T. (1952). Classification and analysis of designs with two associate classes, Journal of Amer. Stat. Assoc., 47, 151-190.

Bose, R.C. and Shrikhande, S.S. (1959a). On the falsity of Euler's conjecture about the non-existence of two orthogonal Latin squares of order 2t + 1. Proc. Natl. Acad. Sc. U.S.A., 45, 734-737.

Bose, R.C. and Shrikhande, S.S. (1959b). On the construction of pairwise orthogonal sets of Latin squares and the falsity of a conjecture of Euler II, Mimeo series No.225, Institute of Statistics, University of North Carolina, Chapell Hill.

Bose, R.C. and Shrikhande, S.S. (1960). On the composition of balanced incomplete block designs, Can. J. Math., 12, 177-188.

Bose, R.C.,Shrikhande,S.S. and Parker, E.T. (1960). Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture, Can. J. Math. 12, 189-203.

Chakrabarti, M.C. (1963). Answer to the query regarding the bounds for common number of treatments between blocks of balanced incomplete block designs, J. Indian Stat. Soc., 1, 230-234.

Cleroux, R. (1969). First and second moments of the randomization test in two associate PBIB designs, J. Amer. Stat. Assoc., 64, 1424-1432.

Cochran, W.G. (1938). Recent works on the analysis of variance, J. Roy. Stat. Soc., 101, 434-449.

Cochran, W.G. (1940). A survey of experimental designs, Mimeo., U.S.A.D.

Cochran, W.G. and Cox, G.M. (1950, 1957). Experimental designs, Wiley, N.Y.

Cornish, E.A. (1940). The estimation of missing values in incomplete randomized block experiments, Ann. Eugen, 10, 112-118.

Das, M.N. and Giri, N. (1976). An alternative method of construction of mutually orthogonal Latin squares of order 3m + 1, (to be published).

De Lury, D.B. (1946). The analysis of Latin squares when some observations are missing, J. Amer. Stat. Assoc., 41, 370-389.

Dembouski, P. (1968). Finite geometries, Springer Verlag.

Federer, W.T. (1963). Experimental designs, Macmillan Co., N.Y.

Fisher, R.A. (1926). The arrangement of field experiments, J. Ministry Agri. 33, 503-513.

Fisher, R.A. (1940). An examination of different possible solutions of a problem in incomplete blocks, Ann. Eugen., 10, 52-75.

Fisher, R.A. (1942). New cyclic solutions of problems in incomplete blocks, Ann. Eugen, 10, 52-75.

Fisher, R.A. (1949). The design of experiments, Oliver Boyd, London.

Fisher, R.A. and Yates, F. (1943,1963). Statistical tables for biological, agricultural and medical research, Oliver Boyd, London.

Giri, N. (1957). On row balance in PBIB designs, J. Indian Soc. Agri.Stat., 11, 168-178.

Giri, N. (1958). On reinforced PBIB designs, J. Indian Soc. Agri. Stat., 12, 41-51.

Giri, N. (1965). On the F-test in the intra-block analysis of a class of partially balanced incomplete block designs, J. Amer. Stat. Assoc., 285-293.

Graybill, F.A. and Weeks, D.L.(1959). Combining inter-block and intra-block information in balanced incomplete blocks, Ann.Math. Math. Statist., 30, 799-805.

Graybill, F.A. and Seshadri, V. (1960). On the unbiasedness of Yate's method of estimation using intra-block information, Ann. Math. Statist. 31, 786-787.

Hartley, H.D., Shrikhande, S.S. and Taylor, W.B. (1953). A note on incomplete block designs with row balance, Ann. Math. Statist. 24, 123-126.

Hotelling, W. (1952). The large sample power of tests based on permutations of observations, Ann. Math. Statist., 23, 169-192.

Kiefer, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs, Ann. Math. Statist. 29, 675-699.

Kirkman, T.P. (1947). On a problem in combinations, Cambridge and Dublin Math. J. 2, 191-204.

Kshirsagar, A.M. (1958). A note on incomplete block designs, Ann. Math. Statist., 29, 907-910.

Lehmann, E. (1959). Testing statistical hypotheses, John Wiley, N.Y.

Ma, R.H. and Harington, H.B. (1948). The standard errors of different designs of field experiments at the University of Saskatchewan, Sci. Agri., 28, 461-474.

Mann, H.B. (1949). Analysis and design of experiments, Dover, New York.

McNeish, H.F. (1922). Eulers squares, Ann. Math., 23, 221-227.

Mitra, S.K. (1968). On the F-test in the intra-block analysis of BIB designs, Sankhya, 22, 279-284.

Mote, V.L. (1958). On a minimax property of a balanced incomplete block designs, Ann. Math. Statist., 29,910-913.

Nair, K.R. (1940). The application of the technique of analysis of covariance to field experiments with several missing or mixed up plots, Sankhya, 4, 581-588.

Nair, K.R. (1944). The recovery of intra-block information in incomplete block designs, Sankhya, 6, 383-390.

Neyman, J. (1923). Sur les applications de la theorie des probabilites aux experiences agricoles, Essay des principles, Roczniki Nauk Rolniczch, 10, 1-51.

Neyman, J. (1935). Statistical problems in agricultural experimentation, J. Royal Stat. Soc., supp. 2, 107-154.

Norton, H.W. (1939). The 7 x 7 squares. Ann. Eng., pp.269-307.

Parker, E.T. (1959). Construction of some sets of orthogonal Latin squares, Proc. Amer. Math. Soc., 10, 946-957.

Rao, C.R. (1947). General methods of analysis for incomplete block designs, J. Amer. Stat. Assoc., 42, 541-561.

Rao, C.R. (1959). Expected values of mean squares in the analysis of incomplete block experiments and some comments based on them, Sankhya 21, 327-336.

Raghava Rao, D.(1971). Construction and combinatorial problems in design of experiments, John Wiley, New York.

Rao, P.V. (1963). The F-test in the intra-block analysis of a class of PBIB designs, Ann. Institute of Stat. Math., 15, 25-36.

Rao, V.R. (1958). A note on balanced designs. Ann. Math. Statist., 29, 290-294.

Rao, S.V.S.P. and Das, M.N. ((1969). Incomplete weighing designs through balanced ternary designs, J. Indian Soc. Agri. Stat. 11, 67-79.

Scheffe, H. (1943). Statistical inference in the nonparametric case, Ann. Math. Statist. 13, 371-388.

Scheffe, H. (1967). The analysis of variance, Wiley, N.Y.

Roy, J. (1958). On the efficiency factor of block designs, Sankhya, 19, 181-188.

Roy, J. and Shah, K.R. (1962). Recovery of inter-block information, Sankhya, A. 24, 269-280.

Shah, K.R. (1964). Use of inter-block information to obtain uniformly better estimators, Ann. Math. Statist. 35, 1064-1075.

Shrikhande, S.S. (1951). Designs for two-way elimination of heterogeneity, Ann. Math. Statist. 22, 235-247.

Sprott, D.A. (1956). A note on combined inter-block and intra-block estimation in incomplete block designs, Ann. Math. Statist. 27, 633-641.

Tocher, K.D. (1952). The design and analysis of block experiments, J. Royal Stat. Soc., B, 14, 45-100.

Vartak, M.N. (1963). Disconnected balanced designs, J. Indian Stat. Assoc., 1, 104-107.

Welch, B.L. (1937). On the Z-test in randomized blocks and Latin squares, Biometrika, 29, 21-52.

Yates, F. (1933). The analysis of replicated experiments when field results are incomplete, Emp. J. Exp. Agri., 1, 129-142.

Yates, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional Lattices, Ann. Eugen., 9, 136-156.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs, Ann. Eugen., 10, 317-325.

Youden, W.J. (1940). Experimental designs to increase accuracy of Greenhouse studies, Cont. Boye Thompson Inst., 11, 219-228.

CHAPTER 6. ANALYSIS OF VARIANCE II, FACTORIAL EXPERIMENTS

This chapter deals mainly with factorial experiments. In Chapter 5 we have discussed analysis of complete and incomplete block designs where the treatments are not necessarily related to each other. In factorial experiments several factors effect simultaneously the characteristic under study and the experimenter is interested in the main effects and interactions between different factors. We shall consider briefly the technique of confounding to get incomplete blocks of smaller sizes when the number of treatments is large. We shall then consider the nested classifications, crossed and nested classifications, split-plot designs and Lattice designs which are some standard incomplete factorial designs commonly used in actual experimentations. The presentation in this chapter is far from complete and our main aim will be to give the analysis of variance of some standard incomplete factorial design. The random-effect model and the mixed-effect model will be introduced as they are very appropriate to analyze the nested classifications, the crossed and nested classifications and the split-plot designs.

We refer to Kempthorne (1951), Federer (1959) and Raghava Rao (1971) for design part of the factorial experiments including both the complete and incomplete ones.

6.1. 2^n-FACTORIAL EXPERIMENTS

A group of treatments which contain two or more levels of two or more factors in all possible combinations is known as a factorial arrangement. For two factors A, B, A with two levels a_0, a_1 and B with two levels b_0, b_1, the four treatment combinations are a_0b_0, a_1b_1, a_0b_1, a_1b_0 .

The effect of the factor A when the factor B is at the level b_0 is $(a_1b_0-a_0b_0)$, and the same when the factor B is at the level b_1 is $(a_1b_1-a_0b_1)$. The average (of these two effects) $\frac{1}{2}(a_1b_1-a_0b_1) + \frac{1}{2}(a_1b_0-a_0b_0)$ is called the main effect of A and is written formally as the expression .

$$(A) = \tfrac{1}{2}(a_1-a_0)\ (b_1+b_0) \tag{6.1}$$

Similarly the main effect (B) of the factor B is given by

$$(B) = \tfrac{1}{2}(b_1-b_0)(a_1+a_0) \tag{6.2}$$

If the two factors A, B act independently with respect to the characteristic under study we would expect

$$a_1b_1-a_0b_1 = a_1b_0-a_0b_0$$

But if they are different the average difference

$$\tfrac{1}{2}(a_1b_1-a_0b_1)-\tfrac{1}{2}(a_1b_0-a_0b_0) = \tfrac{1}{2}(a_1-a_0)(b_1-b_0) = (AB) \tag{6.3}$$

is a measure of the extent to which the factors interact and is called the interaction (AB) between the factors A, B.

For three factors A, B, C, A with levels a_0, a_1; B with levels b_0, b_1 and C with levels c_0, c_1, to find the main effect of A we take the average of the four effects

$$a_1b_0c_0-a_0b_0c_0,\ a_1b_1c_0-a_0b_1c_0,\ a_1b_0c_1-a_0b_0c_1,\ a_1b_1c_1-a_0b_1c_1$$

of the factor A corresponding to each of the four combinations $b_0c_0, b_1c_0, b_0c_1, b_1c_1$ of the levels of the factors B and C. This is written formally as the expression

$$(A) = \tfrac{1}{4}(a_1b_0c_0-a_0b_0c_0+a_1b_1c_0-a_0b_1c_0+a_1b_0c_1-a_0b_0c_1+a_1b_1c_1-a_0b_1c_1) \tag{6.4}$$

$$= \tfrac{1}{4}\ (a_1-a_0)(b_1+b_0)(c_1+c_0)$$

Similarly the main effects of the factors B, and C are given by (6.5)

$$(B) = \tfrac{1}{4}\ (a_1+a_0)(b_1-b_0)(c_1+c_0),$$

$$(C) = \tfrac{1}{4}\ (a_1+a_0)(b_1+b_0)(c_1-c_0)$$

To find the two-factor interaction (AB) of the two factors A and B we find the average effect of the factor A over the levels of the factor C for each level of the factor B, that is, we find

$$\tfrac{1}{2}(a_1b_0c_0-a_0b_0c_0+a_0b_0c_1-a_0b_0c_1) \tag{6.5}$$

$$\tfrac{1}{2}(a_1b_1c_1-a_0b_1c_1+a_1b_1c_0-a_0b_1c_0),$$

These two effects will in general be different and the average difference is called the interaction (AB) and is written formally as

$$(AB) = \tfrac{1}{4}(a_1-a_0)(b_1-b_0)(c_1+c_0) \tag{6.6}$$

Similarly the other two 2-factor interactions are

$$(AC) = \tfrac{1}{4}(a_1-a_0)(b_1+b_0)(c_1-c_0), \quad (BC) = \tfrac{1}{4}(a_1+a_0)(b_1-b_0)(c_1-c_0) \tag{6.7}$$

If we evaluate the interaction (AB) for each level of C we get $\tfrac{1}{2}(a_1b_1c_1+a_0b_0c_1-a_1b_0c_1-a_0b_1c_1)$ when the factor C is at the level c_1 and $\tfrac{1}{2}(a_1b_1c_0+a_0b_0c_0-a_1b_0c_0-a_0b_1c_0)$ when the factor C is at level Obviously the average of these two quantities is the interaction (AB). If these two quantities are different the average difference, written symbolically as

$$(ABC) = \tfrac{1}{4}(a_1-a_0)(b_1-b_0)(c_1-c_0)$$

is called the 3-factor interaction (ABC). Writing

$$\mu = \frac{1}{8}(a_0b_0c_0+a_1b_0c_0+a_0b_1c_0+a_1b_1c_0+a_0b_0c_1+a_1b_0c_1+a_0b_1c_1+a_1b_1c_1),$$

the coefficient of treatments in the linear combinations representing different effects and μ are given in Table 6.1.

Table 6.1. Coefficients of treatments in different effects

Effects	Treatments $a_0b_0c_0$	$a_1b_0c_0$	$a_0b_1c_0$	$a_1b_1c_0$	$a_0b_0c_1$	$a_1b_0c_1$	$a_0b_1c_1$	$a_1b_1c_1$
8μ	1	1	1	1	1	1	1	1
4(A)	-1	1	-1	1	-1	1	-1	1
4(B)	-1	-1	1	1	-1	-1	1	1
4(AB)	1	-1	-1	1	1	-1	-1	1
4(C)	-1	-1	-1	-1	1	1	1	1
4(AC)	1	-1	1	-1	-1	1	-1	1
4(BC)	1	1	-1	-1	-1	-1	1	1
4(ABC)	-1	1	1	-1	1	-1	-1	1

Note. The term "effect" will be used to indicate main effect or interaction of any order. From Table 6.1, it is easy to see that linear combinations representing various effects are mutually orthogonal.

Given n factors each at two levels there are n main effects, $\binom{n}{2}$ 2-factor interactions, $\binom{n}{3}$ 3-factor interactions and so on.

Hence the total number of effects is 2^n-1, which is also the number of degrees of freedoms associated with treatment sum of squares.

The development of factorial experiments is mainly due to the pioneering works of Yates and the reader is referred to Yates (1937) for the treatment of this topic and relevant original references.

6.2. GENERAL FACTORIAL EXPERIMENTS

Let the treatments consist of the combination of n factors $F_1,\dots,F_n$, where the factor F_i is at the s_i levels, $i = 1,\dots,n$. A typical treatment combination in which the i factor F_i is at the level x_i, $i = 1,\dots,n$, may be written as

$$\prod_{i=1}^{n} f_i^{x_i} \quad \text{or} \quad \tau(x_1,\dots,x_n)$$

A linear function of treatments

$$\Sigma c(x_1,\dots,x_n) \quad \tau(x_1,\dots,x_n)$$

is called a contrast if the $c(x_1,\dots,x_n)$, for different values of $x_1,\dots,x_n$, are not all equal to zero and $\Sigma\ c(x_1,\dots,x_n) = 0$, where the summation is taken, over all values of $x_1,\dots,x_n$.

A treatment contrast is said to belong to a k-factor interaction $(F_{i_1}\dots F_{i_k})$ of the factors $F_{i_1},\dots,F_{i_k}$ if for all different values of $x_1,\dots,x_n$

(i) all $c(x_1,\dots,x_n)$ depends only on $x_{i_1},\dots,x_{i_k}$;

(ii) $\underset{x_j}{\Sigma}\ c(x_1,\dots,x_n) = 0$, $j = i_1,\dots,i_k$.

A one-factor interaction is called a main effect. Let ϕ_i denote the sum of all levels of the factor F_i, that is,

$$\phi_i = f_i^1 + \dots + f_i^{s_i} \ , \quad i = 1,\dots,n. \tag{6.8}$$

Then a typical treatment contrast belonging to the k-factor interaction $(F_{i_1}\dots F_{i_k})$ is

$$\frac{\phi_1\cdots\phi_n}{\phi_{i_1}\cdots\phi_{i_k}}\sum \ell(x_{i_1},\ldots,x_{i_k})\, f_{i_1}^{x_{i_1}}\cdots f_{i_k}^{x_{i_k}} \tag{6.9}$$

where the $\ell(x_{i_1},\ldots,x_{i_k})$ for different values of $x_{i_1},\ldots,x_{i_k}$ are not all zero and

$$\sum_{x_j} \ell(x_{i_1},\ldots,x_{i_k}) = 0,\ j = i_1,\ldots,i_k\ . \tag{6.10}$$

One obvious way of obtaining these contrasts is to choose

$$\ell(x_{i_1},\ldots,x_{i_k}) = \prod_{j=i_1}^{i_k} \ell_j(x_j) \tag{6.11}$$

where $\sum_{x_j} \ell_j(x_j)=0$, $j = i_1,\ldots,i_k$. From (6.11) it follows that there are

$\prod_{j=i_1}^{i_k} (s_j-1)$ independent contrasts

$$\frac{\phi_1\cdots\phi_n}{\phi_{i_1}\cdots\phi_{i_k}} \sum_{j=i_1}^{i_k} \ell_j(x_j)\, f_j^{x_j}$$

all belonging to the k-factor interaction $(F_{i_1}\cdots F_{i_k})$. Furthermore any two different contrasts belonging to the same or different interactions and main effects are mutually orthogonal. Thus there are (s_i-1), $i = 1,\ldots,n$, sets of independent (linearly) treatment contrasts all belonging to the main effects and $(s_i-1)(s_j-1)$, $i \neq j$, $i,j = 1,\ldots,n$ sets of independent contrasts all belonging to the 2-factor interactions,...,

$\prod_{i=1}^{n} (s_i-1)$ independent contrasts all belonging to the n-factor interaction. Since

$$\sum_{i=1}^{n} (s_i-1) + \sum_{i\neq j}\sum (s_i-1)(s_j-1) + \cdots + \prod_{i=1}^{n} (s_i-1)$$

$$= \Big[\prod_{i=1}^{n} (1+s_i-1)\Big] - 1 = \prod_{i=1}^{n} s_i - 1\ ,$$

we conclude that the total number of independent contrasts, each with one degree of freedom $\prod_{i=1}^{n} s_i-1$. A factorial experiment (with n-factors) is called symmetrical if $s_1 = \dots = s_n = s$ and is then called a s^n factorial experiment. Then if all factors are of 2 levels, each main effect or interaction of any order has only one degree of freedom.

Example 6.1. Consider two factors A and B, A with levels a_0, a_1, a_2 and B with levels b_0, b_1, b_2. The treatment combinations are $a_0b_0, a_0b_1, a_0b_2, a_1b_0, a_1b_1, a_1b_2, a_2, b_0, a_2b_1, a_2b_2$. The contrasts belonging to (A) (each with one degree of freedom) are

$$(a_2-a_0)\ (b_0+b_1+b_2)\ ,\ (a_2-2a_1+a_0)\ (b_0+b_1+b_2)\ .$$

Similarly the contrasts belonging to (B) are

$$(b_2-b_0)\ (a_0+a_1+a_2),\ (b_2-2b_1+b_0)\ (a_0+a_1+a_2)\ .$$

The contrasts belonging to (AB) are

$$(a_2-a_0)\ (b_2-b_0),\quad (a_2-2a_1+a_0)\ (b_2-b_0)\ ,$$
$$(a_2-a_0)\ (b_2-2b_1+b_0),\ (a_2-2a_1+a_0)\ (b_2-2b_1+b_0)\ .$$

If the factorial experiment can be carried out in a complete block design such as randomized block or latin square design with r replications per treatment the best estimate of any treatment contrast is obtained from the treatment contrast by replacing the treatments by their corresponding average yields as obtained from the experiment. Let $T_{i_1,\dots,i_k}(x_{i_1},\dots,x_{i_k})$ denote the total yield of all treatment combinations where the factor F_j is at the level x_j, $j = i_1,\dots,i_k$. The sum of squares due to the main effect of the factor F_1 is given by

$$SS(F_1) = \sum_{x_1} \frac{T_1^2(x_1)}{r\, s_2\cdots s_n} - \frac{T^2}{r\, s_1 \cdots s_n} \tag{6.12}$$

with s_1-1 degrees of freedom where $T = \sum_{x_1=1}^{s_1} T_1(x_1)$

and similar expressions hold for the main effects of other factors. The sum of squares due to the interaction (F_1F_2) is

$$SS(F_1F_2) = \sum_{x_1}\sum_{x_2} \frac{T_{12}(x_1,x_2)}{r\, s_3\cdots s_n} - \frac{T^2}{r\, s_1\cdots s_n}\ . \tag{6.13}$$

and similar expression holds for any two-factor interaction. It may be noted that $SS(F_1F_2)$ is the error sum of squares of a two-way classification with factors F_1, F_2 and with a single observation $T_{12}(x_1,x_2)$ per cell. Similarly the sum of squares due to interaction $(F_1F_2F_3)$ is given by

$$SS(F_1F_2F_3) = \Big[\sum_{x_1} \sum_{x_2} \sum_{x_3} T^2_{123}(x_1,x_2,x_3) - \frac{T^2}{r\, s_1 \ldots s_n} \Big]$$

$$- SS(F_1) - SS(F_2) - SS(F_1F_2) - SS(F_3)$$

$$- SS(F_1F_3) - SS(F_2F_3). \qquad (6.14)$$

The calculation of sum of squares of other interactions are now obvious. The analysis of variance table of a complete block factorial experiment is given by the analysis of variance table of a randomized block or latin square design depending on the design actually used to carry out the experiment. The only difference in this case is to divide the total sum of squares into the sums of squares of different main effects and interactions. As an example we give the analysis of variance table of a 3-factor factorial experiment, carried out in a randomized block design with r replications per treatment (Table 6.2). The significance of each main effect and interaction is tested by F-test. For example, to test the significance of the main effect (F_1) we use the statistic $F = MSS(F_1)/MSS(E)$ which has F-distribution (s_1-1), $(r-1)(s_1s_2s_3-1)$ degrees of freedom under the hypothesis. Similarly the significance of the interaction (F_1F_2) is tested by the statistic $F = MSS(F_1F_2)/MSS(E)$ which has F distribution with $(s_1-1)(s_2-1)$, $(r-1)(s_1s_2s_3-1)$ degrees of freedom under the hypothesis.

6.3. CONFOUNDING

When a large number of treatment combinations are to be experimentally tested, it is undesirable to use complete block designs because in that case plots in each block may not be homogeneous in their soil fertility. Since this may result in larger experimental error variance it is desirable to keep the

Table 6.2. ANOVA Table of a 3-factor factorial experiment carried out in randomized block designs

Sources of variation	d.f.	Sum of squares	Mean sum of squares	F
Replications	$r-1$	-	-	-
Treatments	$s_1 s_2 s_3-1$	-	-	-
Main effects				
(F_1)	s_1-1	$SS(F_1)$	$MSS(F_1)$	$\frac{MSS(F_1)}{MSS(E)}$
(F_2)	s_2-1	$SS(F_2)$	$MSS(F_2)$	$\frac{MSS(F_2)}{MSS(E)}$
(F_3)	s_3-1	$SS(F_3)$	$MSS(F_3)$	$\frac{MSS(F_3)}{MSS(E)}$
Interactions				
(F_1F_2)	$(s_1-1)(s_2-1)$	$SS(F_1F_2)$	$MSS(F_1F_2)$	$\frac{MSS(F_1F_2)}{MSS(E)}$
(F_1F_3)	$(s_1-1)(s_3-1)$	$SS(F_1F_3)$	$MSS(F_1F_3)$	$\frac{MSS(F_1F_3)}{MSS(E)}$
(F_2F_3)	$(s_2-1)(s_3-1)$	$SS(F_2F_3)$	$MSS(F_2F_3)$	$\frac{MSS(F_2F_3)}{MSS(E)}$
$(F_1F_2F_3)$	$(s_1-1)(s_2-1)(s_3-1)$	$SS(F_1F_3F_3)$	$MSS(F_1F_2F_3)$	$\frac{MSS(F_1F_2F_3)}{MSS(E)}$
Error	$(r-1)(s_1s_2s_3-1)$	$SS(E)$	$MSS(E)$	
Total	$rs_1s_2s_3-1$	-	-	

block size small. By a device known as "confounding" the necessity of all treatments in each block in the randomized block design or in each row and column in a latin square design is side-stepped (see Yates(1933,1935,1937)) and the whole block or the replicate is divided into a desired number of incomplete blocks. To understand the concept of confounding let us consider the following example.

Example 6.2. Consider a factorial experiment with 3 factors A, B and C, A with levels a_0, a_1, B with levels b_0, b_1 and C with levels c_0, c_1. There are 8 treatment combinations, for a complete factorial randomized block design each block contains all the 8 treatment combinations. Consider the interaction

$$\begin{aligned}(ABC) &= \tfrac{1}{4}(a_1-a_0)(b_1-b_0)(c_1-c_0)\\ &= \tfrac{1}{4}(a_1b_1c_1+a_1b_0c_0+a_0b_1c_0+a_0b_0c_1)\\ &\quad - \tfrac{1}{4}(a_1b_1c_0+a_1b_0c_1+a_0b_1c_1+a_0b_0c_0)\end{aligned}$$

and suppose that we arrange 8 treatment combinations in the following 2 blocks (the figure inside the brackets denotes the treatment number).

Block 1	Block 2
$a_1b_1c_1$ $a_1b_0c_0$ $a_0b_0c_0$ $a_0b_0c_1$,	$a_1b_1c_0$ $a_1b_0c_1$ $a_0b_1c_1$ $a_0b_0c_0$.

The block-1 contains all treatment combinations with positive sign and the block-2 contains all treatment combinations with negative sign in the expression for (ABC). Let x_{ij} denote the observation on the jth treatment in the ith block and let

$$x_{ij} = \mu + \alpha_i + t_j + e_{ij},\ i = 1,2,\ j = 1,\dots,8,$$

where μ is the grand mean, α_i is the ith block effect satisfying $\alpha_1 + \alpha_2 = 0$, t_j is the jth treatment effect satisfying $\sum\limits_j t_j = 0$ and the e_{ij} are observation errors. Let $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{t}_j$ denote the estimate of μ, α_i, t_j respectively. Then

$$\begin{aligned}x_{18} &= \hat{\mu} + \hat{\alpha}_1 + \hat{t}_8\,, & x_{24} &= \hat{\mu} + \hat{\alpha}_2 + \hat{t}_4\\ x_{17} &= \hat{\mu} + \hat{\alpha}_1 + \hat{t}_7\,, & x_{23} &= \hat{\mu} + \hat{\alpha}_2 + \hat{t}_3\\ x_{16} &= \hat{\mu} + \hat{\alpha}_1 + \hat{t}_6\,, & x_{22} &= \hat{\mu} + \hat{\alpha}_2 + \hat{t}_2\\ x_{15} &= \hat{\mu} + \hat{\alpha}_1 + \hat{t}_5\,, & x_{21} &= \hat{\mu} + \hat{\alpha}_2 + \hat{t}_1.\end{aligned}$$

Consider

$$\begin{aligned}&\tfrac{1}{4}(x_{18} + x_{17} + x_{16} + x_{15} - x_{24} - x_{23} - x_{22} - x_{21})\\ &= \tfrac{1}{4}(\hat{t}_8 + \hat{t}_7 + \hat{t}_6 + \hat{t}_5 - \hat{t}_4 - \hat{t}_3 - \hat{t}_2 - \hat{t}_1) + \tfrac{1}{2}(\hat{\alpha}_1 - \hat{\alpha}_2)\\ &= \tfrac{1}{4}(ABC) + \tfrac{1}{2}(\hat{\alpha}_1 - \hat{\alpha}_2)\,.\end{aligned}$$

From this it is obvious that it is not possible to estimate (ABC) free from block effects. In other words the interaction (ABC) and block differences are extricably mixed up or completely confounded with each other. It is worth noting that if we have one complete block instead of two incomplete blocks then (ABC) can be estimated free from the block effects (here $\alpha_1 = \alpha_2$).

With the incomplete blocks, 1, 2, the other main effects and interactions are not confounded with the block differences. To see this consider the main effect (A)

$$(A) = \tfrac{1}{4}(a_1b_1c_1 + a_1b_0c_1 + a_1b_1c_0 + a_1b_0c_0 - a_0b_1c_1 - a_0b_0c_1 - a_0b_1c_0 - a_0b_0c_0) .$$

Since

$$\tfrac{1}{4}(x_{18}+x_{23}+x_{24}+x_{17}-x_{22}-x_{15}-x_{16}-x_{21})$$

$$= \tfrac{1}{4}(\hat{t}_8 + \hat{t}_3 + \hat{t}_4 + \hat{t}_7 - \hat{t}_2 - \hat{t}_5 - \hat{t}_6 - \hat{t}_1),$$

(A) is not confounded with the block differences. Similar conclusions hold for other effects except (ABC).

In Example 6.2 we have considered only a single replication. However, we could have r replications by replicating those two incomplete blocks r times or by taking different pairs of incomplete blocks for different replications. To get a different pair of incomplete blocks of size 4 we obviously need to confound a main effect or a 2-factor interaction with the block difference. By replicating the same pair of incomplete blocks r times we confound (ABC) completely with the block difference in the whole experiment.

If a main effect or an interaction is of little or no importance, it is recommended to confound it with block differences and such an effect is called "completely confounded". If an effect is confounded with block differences in some replicates but is not confounded in the remaining replicates, it is called partially confounded.

To reduce further the block sizes we need to confound two or more effects simultaneously. Suppose in the above example we want incomplete blocks of size 2 and we decide to confound (ABC) and (A). Assigning all treatment combinations having "+" sign in both (ABC) and (A) in block 1, treatment combinations having plus sign in (ABC) and minus sign in (A) in block 2, treatment combinations having minus sign in (ABC) and + sign in (A) in block 3 and finally treat combinations having - sign in both (AB) and (A) in block 4. We obtain the following arrangement in blocks of size 2.

Block 1	Block 2	Block 3	Block 4
$a_1b_1c_1$ $a_1b_0c_0$	$a_0b_1c_0$ $a_0b_0c_1$	$a_0b_0c_0$ $a_0b_1c_1$	$a_0b_0c_1$ $a_0b_1c_0$

In such confounding arrangement, in addition to (ABC) and (A) which we have confounded to get these blocks, the interaction (BC) is automatically confounded. The interaction (BC) is called the generalized effect of (ABC) and (A). In general the generalized effect of any two effects is a treatment contrast wherein the coefficient of a particular treatment combination is 1 if that treatment combination occurs with the same sign in both effects and is -1 otherwise. It is easily verified that the generalized effect of any two effects (X) and (Y) where X and Y contain different selections of the factors A,B,C, etc. is obtained by multiplying the factors in X,Y together and equating the square of any factor to unity. For example

$$(A)\ (ABC) = (A^2BC) = (BC), \quad (AB)(BC) = (AB^2C) = (AC).$$

The general rule exemplified by this case is that if any two effects are completely confounded with block differences so is their generalized effect. The analysis and constructions of confounded, partially confounded factorial experiments are well treated in Kempthorne (1951), Federer (1955), Johns (1971) Raghava Rao (1971), Das (1964) and Das and Rao (1967). In the following sections we shall discuss the analysis of some standard incomplete factorial designs which are widely used in scientific experimentation.

6.4. NESTED DESIGNS (FIXED EFFECT MODEL)

Here we shall consider treatments which contain two or more levels of two or more factors and we are interested to find out the relationship of factors in an incomplete layout called nesting. The levels of a factor B are said to be nested within the levels of a factor A or simply B is nested in A if every level of the factor B occurs with only a single level of the factor A. We call two factors A, B crossed if neither is nested in the other and completely crossed if every level of one occurs with every level of the others, that is, they form a complete two-way layout.

Consider a genetical experiment with I sires where each of the I sires is mated with J dams and n daughters are chosen from each of the IJ litters and we study the variability of the lactation yields of daughters. This is an example of a nested design with two factors A (sires), B(dams) and B is nested in A. Let x_{ijk} denote the lactation yield of the kth daughter of the jth dam mated with the ith sire. Though random effect or mixed effect models are more realistic for nested designs let us first consider the analysis of the fixed-effect model for nested designs. The fixed-effect model of this experiment is

$$x_{ijk} = \mu + \alpha_i + t_j + e_{ijk} \qquad (6.18)$$

$i = 1,\ldots,I$; $j = 1,\ldots,J$; $k = 1,\ldots,n$ where μ is the grand mean, α_i is the effect of the ith sire, t_{ij} is the effect of the jth dam mated with the ith sire and the e_{ijk} is the effect of the kth daughter of the jth dam mated with the ith sire. In the fixed effect model the e_{ijk} are observational errors which are assumed to be independently normally distributed with means zero and the same variance σ^2. We also assume that $\sum_i \alpha_i = 0$, $\sum_j t_{ij} = 0$ for each i.

It is not necessary that the number of dams mated with different sires be the same, they may vary from sire to sire. Similarly the number of daughters on different dams may also vary. Let the number of daughters on the jth dam mated with ith sire be n_{ij} and the number of dams mated with the ith sire be J_i. Write

$$E(X_{ijk}) = \mu_{ij} = \mu + \alpha_i + t_j$$

$$x_{ijk} = \mu_{ij} + e_{ijk},$$

$$n_i = \sum_{j=1}^{J_i} n_{ij}, \quad n = \sum_{i=1}^{I} n_i, \tag{6.19}$$

$k = 1,\ldots,n_{ij}$, $j = 1,\ldots,J_i$, $i = 1,\ldots,I$. Suppose we impose the restriction that

$$\sum_i \frac{n_i\alpha_i}{n} = 0 \quad \sum_j \frac{n_{ij}t_{ij}}{n_i} = 0 \quad \text{for all } i\ . \tag{6.20}$$

Then with $\mu_{ij} = \mu + \alpha_i + t_{ij}$ we obtain

$$\sum_j \frac{n_{ij}\mu_{ij}}{n_i} = \mu + \alpha_i, \qquad \sum_i \sum_j \frac{n_{ij}\mu_{ij}}{n} = \mu \tag{6.21}$$

The fixed effect model for the general 2-factor nested design is then

$$x_{ijk} = \mu + \alpha_i + t_j + e_{ijk} \tag{6.22}$$

where μ , α_i, t_{ij} are as defined in (6.21) subject to restrictions (6.20) and the e_{ijk} are independently normally distributed with means 0 and the same variance σ^2. With this model we are generally interested in testing the hypothesis $H_{\underline{\alpha}}$: all $\alpha_i = 0$, H_T: all $t_{ij} = 0$. To obtain the error sum of squares (to estimate σ^2) we have to minimize

$$s^2 = \sum_i \sum_j \sum_k (x_{ij} - \mu_{ij})^2 \tag{6.23}$$

with respect to μ_{ij}. The least square estimates of the μ_{ij} are evidently $\hat{\mu}_{ij} = x_{ij.} = \sum_k x_{ijk}/n_{ij}$ and the error sum of squares is

$$\sum_i \sum_j \sum_k (x_{ijk} - \hat{\mu}_{ij})^2 = \sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2, \tag{6.24}$$

and is easily seen to have $\sum_i \sum_j (n_{ij}-1)$ degrees of freedom.

Under $H_{\underline{\alpha}}$ the minimum value of s^2 is

$$\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2 + \sum_i n_i x_{i..}^2 - n x^2_{\ldots} \tag{6.25}$$

and under H_T the minimum value of s^2 is

$$\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2 + \sum_i \sum_j n_{ij} x_{ij.}^2 - \sum_i n_i x_{i..}^2 \qquad (6.26)$$

where

$$x_{...} = \sum_i \sum_j \sum_k \frac{x_{ijk}}{n}, \quad x_{i..} = \sum_j \sum_k \frac{x_{ijk}}{n_i} .$$

Thus from (6.24) and (6.26) the appropriate invariant test of H_T (at level α) rejects H_T whenever

$$\frac{d_3 q_2}{d_2 q_3} \geq F_{1-\alpha, d_2, d_3} \qquad (6.27)$$

where d_2, d_3, q_2 and q_3 are as given in Table 6.3. Similarly from (6.24) and (6.25) with the additive model (6.18), one could use for testing $H_{\underline{\alpha}}$ the invariant test which rejects $H_{\underline{\alpha}}$ whenever

$$\frac{d_3 q_1}{(I-1) q_3} \geq F_{1-\alpha, I-1, d_3} \qquad (6.28)$$

where q_1 is as given in Table 6.3. However a more appropriate invariant test of $H_{\underline{\alpha}}$, which is valid whether the assumption of additivity holds or not, rejects $H_{\underline{\alpha}}$ whenever

$$\frac{d_2 q_1}{(I-1) q_2} \geq F_{1-\alpha, I-1, d_2} \quad . \qquad (6.29)$$

However, when all $t_{ij} = 0$ (6.28) and (6.29) are the same.

Table 6.3. ANOVA table of a two-factor completely nested classification with fixed effect model

Sources of variation	d.f.	Sum of squares	F
Sires	$I-1$	$q_1 = \sum_i n_i x_{i..}^2 - n x_{...}^2$	$\frac{d_2 q_1}{(I-1) q_2}$
Dames within sires	$d_2 = \sum_i (J_i - 1)$	$q_2 = \sum_i \sum_j n_{ij} x_{ij.}^2 - \sum_i n_i x_{i..}^2$	$\frac{d_3 q_2}{d_2 q_3}$
Error	$d_3 = \sum_i \sum_j (n_{ij} - 1)$	$q_3 = \sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2$	-
Total	$n-1$	$\sum_i \sum_j \sum_k (x_{ijk} - x_{...})^2$	-

As for nested designs the random effect model is more realistic than the fixed effect model, the F statistics given in Table 6.3 for the fixed effect model have not got much practical interest. We shall consider the analysis of a nested design with random-effect model in Section 6.5 after we discuss the general concept of random-effect models in the following section.

6.5. RANDOM-EFFECT MODEL

Let us consider the one-way classification with linear model (see (3.94)).

$$E(X_{ij}) = \mu + a_i, \quad \text{or} \quad x_{ij} = \mu + a_i + e_{ij}, \; j = 1,\ldots,N_i, \; i=1,\ldots,p. \tag{6.30}$$

If we suppose that the a_i is a random sample of size p (instead of unknown constants as in the fixed-effect model) from a normal population with mean 0 and variance σ_A^2 and the e_{ij} are independently normally distributed with mean 0 and variance σ^2 and are independent of the a_i then (6.30) becomes the random-effect model of the one-way classification. Since from (6.30) $\mathrm{Var}(X_{ij}) = \sigma_A^2 + \sigma^2$, σ_A^2 and σ^2 are components of the variance of X_{ij} and are called variance components. In the fixed-effect model we write $a_i = \alpha_i$.

Example 6.3. In an experiment to study the variation of lactation yield of daughters of different dams, a genetist might pick p dams at random from the population under study and then choose n_i daughters from the ith dam selected for the experiment. In such a case x_{ij} will denote the lactation yield of jth daughter of the ith dam. We assume that this gives a mean lactation yield μ, σ_A^2 denote the dam-to-dam variation and σ^2 measures to variability of daughters' yield from the same dam.

Example 6.4. Suppose that an experiment is tried in a factory with p workers, selected randomly from a large number of workers who produce small parts of some kind. Suppose also that a large number of parts is produced daily on the machine and for any worker there is considerable day-to-day variation.

Suppose that the ith worker is assigned for n_i days during the experiment. Denoting by x_{ij} the output of the ith worker in the jth day that he uses the machine we can write

$$x_{ij} = \mu + \alpha_i + e_{ij}, \quad j = 1,\ldots,N_i, \quad i = 1,\ldots,p,$$

where $\mu + \alpha_i$ is the true mean output of the ith worker, an idealized daily average for the ith worker after he has acquired the stability with the machine following a learning stage and e_{ij} is his error on the jth day. The variability of outputs of the ith worker about the true mean $\mu + \alpha_i$ is measured by σ_A^2 (assuming that they are the same for all workers in the population).

The hypothesis usually tested in a random effect model of a one-way classification is $H_{\underline{a}}:\sigma_A^2 = 0$. Obviously $H_{\underline{\alpha}}$ is true if and only if all a_i are equal. In the one-way classification if all N_i are equal, then we call it a balanced classification. Generally a higher-way classification with several factors is called balanced if the number of observations in the cells are all equal. A layout with nested factors is called balanced if the number of levels of a nested factor is the same for every combination of all other factors within which it is nested, the factors (if any) which are crossed, are completely crossed; and if the number of replications is the same for every combination of other factors in the experiment.

We shall now consider the analysis of a balanced one-way classification with random-effect model $(N_1 = \cdots = N_p = n)$. Wald (1940) solved this problem for the general case of unequal N_i but his solutions are not very convenient for practical computations.

Using the notation of Table 3.4 the maximum likelihood estimate of the components of variance are given by

$$\hat{\sigma}^2 = [p(n-1)]^{-1} q_2, \quad n\hat{\sigma}_A^2 + \hat{\sigma}^2 = (p-1)^{-1} q_1 \tag{6.31}$$

In a one-way classification it is relatively simple to find these maximum likelihood estimates. However, in more complex experiments with random effect model the solution to the maximum likelihood equations are not so easily obtained. The general approach, to obtain tests and estimates in random-effect models in balanced cases, is to consider the mean squares in the corresponding analysis of variance table under the fixed-effect model. It may be added that the statistics usually suggested with the random-effect models are not necessarily the ones used with the fixed-effect models.

To explain this fully let us first find out the distribution of

$$Q_1 = \sum_i n(X_{i.}-X_{..})^2, \qquad Q_2 = \sum_i \sum_j (X_{ij}-X_{i.})^2 \tag{6.32}$$

for the balanced one-way classification with random-effect model. In the random-effect model (as also we have observed in Chapter 5 in connection with the randomization model) $p^{-1} \sum_i a_i = a_.$ cannot be taken to be equal to zero. Since the e_{ij} are independently normally distributed with mean zero and variance σ^2

$$\frac{Q_2}{\sigma^2} = \sum_i \sum_j \frac{(e_{ij}-e_{i.})^2}{\sigma^2} \tag{6.33}$$

is distributed as the central chisquare with $p(n-1)$ degrees of freedom. Now

$$Q_1 = \sum_i n(a_i-a_. + e_{i.}-e_{..})^2 = \sum_i n(a_i + e_{i.} - a_.-e_{..})^2 \tag{6.34}$$

From the independence of the a_i and e_{ij} and the fact that the a_i are independently normally distributed with mean 0 and variance σ_A^2, $\text{Var}\,(a_i + e_{i.}) = \sigma_A^2 + \frac{\sigma^2}{n}$. Thus

$$\frac{Q_1}{n\,\sigma_A^2+\sigma^2} = \sum_i \frac{(a_i + e_{i.} - a_.-e_{..})^2}{\sigma_A^2 + n^{-1}\,\sigma^2} \tag{6.35}$$

is distributed as χ^2_{p-1}.

Obviously from our assumption in the random-effect model the $(a_i-a_.)$ and the $(e_{ij}-e_{i.})$ are independent. From the assumption of independence of the e_{ij}

$$\text{Cov}(e_{ij}-e_{i.}\,,\, e_{i.}-e_{..}) = E((e_{ij}-e_{i.})e_{i.}) - E(e_{..}(e_{ij}-e_{i.}))$$

$$= E((e_{ij}-e_{i.})e_{i.}) - E(p^{-1}(e_{ij}-e_{i.})e_{i.}) = 0$$

as

$$E((e_{ij}-e_{i.})e_{i.}) = \frac{n-1}{n^2} E(e_{ij}^2) - \frac{1}{n^2} \sum_{j'\neq j} e_{ij'}^2$$

$$= \frac{n-1}{n^2}\sigma^2 - \frac{n-1}{n^2} = 0 .$$

Thus $a_i - a_. + e_{i.} - e_{..}$ and $e_{ij} - e_{i.}$ are independent and hence Q_1, Q_2 are independent in the random-effect model. From (6.33) and (6.35), to test the hypothesis $H_{\underline{a}}$ we use the statistic

$$\frac{p(n-1)}{(p-1)} \frac{Q_1}{Q_2} = (1 + n\frac{\sigma_A^2}{\sigma^2}) F \tag{6.36}$$

where F has central F-distribution with p-1,n(p-1) degrees of freedom and under $H_{\underline{a}}$ $(1 + \frac{n\sigma_A^2}{\sigma^2})F = F$. Thus the level α test of $H_{\underline{a}}$ consists of rejecting $H_{\underline{a}}$ if

$$\frac{p(n-1)q_1}{(p-1)q_2} \geq F_{1-\alpha,p-1,p(n-1)}. \tag{6.37}$$

The power of this test depends on $\theta = \sigma_A^2/\sigma^2$ and is given by

$$\beta(\theta) = P\{(1 + n\theta)\ F \geq F_{1-\alpha,p-1,p(n-1)}\}$$

$$= P\{F \geq (1 + n\theta)^{-1}\ F_{1-\alpha,p-1,p(n-1)}\}.$$

and it involves only central F-distribution.

No doubt the hypothesis $H_{\underline{a}}: \sigma_A^2 = 0$ is not a realistic one but it can be treated as the limiting case of a more realistic hypothesis $H(\theta_0): \theta \leq \theta_0$, where $\theta_0 \geq 0$ is a preassigned constant. The hypotheses $H_{\underline{a}}$, $H(\theta_0)$ correspond respectively to the fixed-effect model hypotheses $\sum_i \alpha_i^2 = 0$, $\sum_i \alpha_i^2/\sigma^2 \leq \theta_0$ and $H_{\underline{a}}$ is a special case of $H(\theta_0)$ when $\theta_0 = 0$. The uniformly most powerful invariant test of $H(\theta_0)$ (see Lehmann (1959), p.288) rejects $H(\theta_0)$ whenever

$$\frac{p(n-1)q_1}{(p-1)\ q_2} \geq C \tag{6.38}$$

where the constant C is determined by the condition that

$$P\{(1 + P\theta)F \geq C\} \leq \alpha \text{ and is equal to } \alpha \text{ if } \theta = \theta_0, \tag{6.39}$$

where α is the level of significance of the test procedure. Thus with $\theta = \theta_0$ (6.39) reduces to

$$P\{(1 + n\theta_0)\, F \geq C\} = \alpha$$

and hence $C = (1 + n\theta_0)F_{1-\alpha,p-1,p(n-1)}$. The power of the test in (6.38) is

$$\beta(\theta) = P\{F \geq \frac{1+n\theta_0}{1+n\theta} F_{1-\alpha,p-1,p(n-1)}\}$$

and it involves only central F-distribution. In the particular case $\theta_0 = 0$ the test (6.37) is equivalent to the fixed-effect model F-test though, of course, they are solutions of different problems and have different power functions.

The uniformly most powerful invariant test of $H(\theta_0): \theta \geq \theta_0$ rejects $H(\theta_0)$ whenever $(1 + n\theta_0)F \leq (1 + n\theta_0)F_{1-\alpha,p-1,p(n-1)}$. The tests of the hypotheses $\theta \leq \theta_0$ and $\theta \geq \theta_0$ are not only uniformly most powerful invariant but also uniformly most powerful unbiased and also the uniformly most powerful unbiased test exists for testing $\theta = \theta_0$ against two-sided alternatives $\theta \neq \theta_0$ (see Lehmann, 1959, p.288). Unbiased estimates $\hat{\sigma}^2, \hat{\sigma}_A^2$ of σ^2, σ_A^2 are obtained from the equations

$$q_2 = p(n-1)\, \hat{\sigma}^2, \quad q_1 = (p-1)(n\, \hat{\sigma}_A^2 + \hat{\sigma}^2). \tag{6.40}$$

The analysis of two-way and higher-way classifications with random-effect models can be similarly treated and we refer to Scheffe (1959) for details.

6.6. NESTED DESIGNS WITH RANDOM-EFFECT MODEL

We shall consider now the analysis of balanced 2-factor nested designs with random effect models. The extension to the general case can be easily carried out by induction with respect to the number of factors. The random-effect model of a two-factor balanced nested design is

$$x_{ijk} = \mu + a_i + t_{ij} + e_{ijk}, \tag{6.41}$$

$i = 1,\ldots,I$, $j = 1,\ldots,J$, $k = 1,\ldots,n$ where the a_i, the t_{ij} and the e_{ijk} are independently normally distributed with means 0 and variances σ_A^2, σ_T^2 and σ^2 respectively. With this model the hypotheses of interest are

$$H_3 : \sigma_A^2 = 0, \quad H_T' : \sigma_T^2 = 0 \tag{6.42}$$

As we have done in the case of one-way classification we have to find first the distribution of (see Table 6.3 for notations)

$$Q_1 = \sum_i nJ(X_{i..} - X_{...})^2, \quad Q_2 = \sum_i \sum_j n(x_{ij.} - X_{i..})^2$$

$$Q_3 = \sum_i \sum_j \sum_k (X_{ijk} - X_{ij.})^2. \tag{6.43}$$

From (6.41) and (6.43)

$$Q_1 = \sum_i nJ(a_i + t_{i.} + e_{i..} - a_{.} - t_{..} - e_{...})^2 \tag{6.44}$$

$$Q_2 = \sum_i \sum_j n(t_{ij} + e_{ij.} - t_{i.} - e_{i..})^2,$$

$$Q_3 = \sum_i \sum_j \sum_k (e_{ijk} - e_{ij.})^2$$

where the dot notation is used to indicate the average with regard to the subscript involved. From (6.44), Q_3/σ^2 is distributed as chisquare with IJ $(n-1)$ degrees of freedom. As the t_{ij} and e_{ijk} are independently normally distributed and

$$E((e_{ijk} - e_{ij.}) = 0, \quad E((e_{ijk} - e_{ij.})e_{i..}) = 0 \tag{6.45}$$

we conclude that Q_2, Q_3 are independently distributed.

Since

$$\mathrm{Var}(t_{ij} + e_{ij.}) = \sigma_T^2 + n^{-1}\sigma^2 \tag{6.46}$$

$Q_2/(n\sigma_T^2+\sigma^2)$ is distributed as central chisquare with $I(J-1)$ degrees of freedom. Furthermore since the a_i, the t_{ij} and the e_{ijk} are independently distributed, the independence of Q_1, Q_2 and that of Q_1, Q_3 follow from (6.45) and the fact that

$$E((t_{ij} - t_{i.})t_{i.}) = 0, \quad E((t_{ij} - t_{i.})t_{..}) = 0,$$

$$E((e_{ij.} - e_{i..})e_{i..}) = 0, \quad E(e_{ij.} - e_{i..})e_{...} = 0,$$

$$E((e_{ijk} - e_{ij.})e_{...} = 0 .$$

Since
$\mathrm{Var}(a_i + t_{i.} + e_{i..}) = \sigma_A^2 + J^{-1}\sigma_T^2 + J^{-1}n^{-1}\sigma^2$, we conclude that $(nJ\sigma_A^2 + n\sigma_T^2 + \sigma^2)^{-1} Q_1$ is distributed as central chisquare with I-1 degrees of freedom. Thus the appropriate level α test of H_T' rejects H_T' whenever

$$\frac{IJ(n-1)q_2}{I(J-1)q_3} \geq F_{1-\alpha, I(J-1), IJ(n-1)}, \tag{6.47}$$

and the level α test of $H_{\underline{a}}$ rejects $H_{\underline{a}}$ whenever

$$\frac{I(J-1)q_1}{(I-1)\ q_2} \geq F_{1-\alpha, I-1, I(J-1)}. \tag{6.48}$$

As in the case of random-effect model of the one-way classification the power of each of these tests can be expressed in terms of central F-distribution.

The hypothesis $H_{\underline{a}}$ and H_T' are not realistic in actual practice but can be expressed as the limiting case where $\theta_0 \to 0$ of the following realistic hypotheses

$$H_{\underline{a}}^* : \frac{\sigma_A^2}{\sigma^2 + n\ \sigma_T^2} \leq \theta_0, \quad H_T^*: \frac{\sigma_T^2}{\sigma^2} \leq \theta_0. \tag{6.49}$$

The uniformly most powerful invariant unbiased level α test of H_T^* (see Lehmann(1959) p.291) whenever

$$\frac{I(J-1)\ q_1}{(I-1)\ \ q_2} \geq (1+n\theta_0)F_{1-\alpha, I-1, I(J-1)} \tag{6.50}$$

and the uniformly most powerful invariant unbiased level α test of $H_{\underline{a}}^*$ rejects $H_{\underline{a}}^*$ whenever

$$\frac{IJ(n-1)q_2}{I(J-1)\ q_3} \geq (1+nJ\theta_0)\ F_{1-\alpha, I(J-1), IJ(n-1)} \tag{6.51}$$

The unbiased estimates $\hat{\sigma}_A^2$, $\hat{\sigma}_T^2$, $\hat{\sigma}^2$ of the variance components σ_A^2, σ_T^2, σ^2 are obtained by solving

$$q_3 = IJ(n-1)\ \hat{\sigma}^2, \quad q_2 = I(J-1)\ (n\ \hat{\sigma}_T^2 + \hat{\sigma}^2)$$
$$q_1 = (I-1)\ (nJ\ \hat{\sigma}_A^2 + n\ \hat{\sigma}_T^2 + \hat{\sigma}^2).$$

6.7. CROSSED AND NESTED DESIGNS (MIXED EFFECT MODEL)

Consider an experiment with three factors A, B, C, A with I levels, B with J levels, C with K levels, in which the factor B is nested within A and the factor C is crossed with factors A and B. As an example consider a chemical process in which a feed stock is pretreated in one of I different ways to form an intermediate and then the intermediate is processed in the next stage at one of K different temperatures. Suppose that we prepare J batches by each of I pretreatment method and from each batch we take Kn samples, n of these are treated at each of K temperatures. In the fixed-effect model the observation $x_{ijk\ell}$, on the ℓ-th sample ($\ell=1,\ldots,n$) which was treated at the kth temperature ($k = 1,\ldots,K$), taken from the jth batch ($j = 1,\ldots,J$) that was prepared by the ith pretreatment method ($i = 1,\ldots,I$), can be written as

$$x_{ijk\ell} = \mu_{ijk} + e_{ijk\ell} \qquad (6.52)$$

with

$$E(X_{ijk\ell}) = \mu_{ijk} = \mu + \alpha_i^A + \alpha_{j(i)}^B + \alpha_k^C + \alpha_{ik}^{AC} + \alpha_{jk(i)}^{BC} \qquad (6.53)$$

where μ is the grand mean, α_i^A is the main effect of the ith level of the factor A, $\alpha_{j(i)}^B$ is the main effect of jth level of the factor B for the ith level of A, α_k^C is the main effect of the kth level of the factor C. Since

$$\alpha_i^A = E(X_{i\ldots}) - \mu, \quad \alpha_{j(i)}^B = E(X_{ij..} - X_{i\ldots}),$$

$$\alpha_k^C = E(X_{..k.} - \mu),$$

We can add restriction $\alpha_{\cdot}^A = \alpha_{\cdot(i)}^B = \alpha_{\cdot}^C = 0$. There are two interaction terms (i) α_{ik}^{AC} which represents the non-additivity between the factor A and the factor C (it may be that changing from one temperature to the other produces a greater change in response to feed stock pretreated by a particular method than those pretreated by others) and (ii) $\alpha_{jk(i)}^{BC}$ having similar interpretation for the ith level of A. Since

$$\alpha^{AC}_{ik} = E(X_{i.k.} - \mu - \alpha^{A}_{i} - \alpha^{C}_{k}) = E(X_{i.k.} - X_{i...} - X_{..k.} + X_{....})$$

we conclude that

$$\alpha^{AC}_{i.} = \alpha^{AC}_{.k} = 0.$$

Similarly

$$\alpha_{.k(i)} = \alpha_{j.(i)} = 0.$$

There is no interaction term between the factors A and B. Since the factor B is nested in A we can no longer talk about the main effect of batches changing from method to method. Two observations made at the same temperature on two different pretreatment stocks have in common one thing that they are made at the same temperature. The first batch made by a particular method and the first batch made by another different method do not have a batch factor in common, as there are just b batches made by each method and being a particular one or the other is just a label. Since the interaction (ABC) corresponds to the interaction AB varying from temperature to temperature, there is no interaction (ABC).

We can analyse this design by treating it as a two-way classification with batches and temperature, the batches being at IJ levels and temperature at K levels and there are n observations per cell. The analysis of variance table for this classification is given in Table 6.4.

We did not write down specifically the null hypotheses involved in obtaining the different sum of squares in Table 6.4. The reader will have no difficulty to figure them out if needed. For example the null hypothesis to obtain SS(C) is all $\alpha^{C}_{k} = 0$, etc. The SS(B) has been divided into two component SS(A), SS(B|A) representing the sum of squares between methods; the sum of squares between batches within methods. Similarly SS(BC) is divided into two components SS(AC) and SS(BC|A), where SS(BC) is the sum of squares for the interaction that we would obtain by ignoring batches and considering a two-way classification with methods (I levels) and temperatures (K levels) and Jn observations per cell and SS(BC|A) is the sum of squares for interaction between temperature and batches made by the same method (within batches).

Table 6.4. Analysis of variance of the three-factors crossed and nested design

Sources of variations	d.f.	Sum of squares	Expected mean sum of squares (under mixed effect model)
Methods (A)	I-1	$SS(A) = \sum_i JKnx^2_{i...} - C.F.$	$\sigma^2 + nK\sigma^2_{B(A)} + JKn\sigma^2_A$
Batches (within methods)	I(J-1)	$SS(B\|A) = SS(B) - SS(A)$	$\sigma^2 + Kn\sigma^2_{B(A)}$
Batches (B)	IJ-1	$SS(A) = \sum_i \sum_j Knx^2_{ij..} - C.F.$	-
Temperatures (C)	K-1	$SS(C) = \sum_k IJnx^2_{..k.} - C.F.$	$\sigma^2 + n\sigma^2_{BC(A)} + IJn\sigma^2_C$
Interaction (AC)	(I-1) (J-1)	$SS(AC) = \sum_i \sum_k Jnx^2_{i.k.} - C.F.$	$\sigma^2 + n\sigma^2_{BC(A)} + IJn\sigma^2_C$
Interaction (BC) within methods	I(J-1) (K-1)	$SS(BC\|A) = SS(BC) - SS(AC)$	-
Interaction (BC)	(IJ -1) (K-1)	$SS(BC) = \sum_i \sum_j \sum_k nx^2_{ijk.} - C.F.$	-
Error	IJK(n-1)	$SS(E) = s^2 - SS(B) - SS(C) - SS(BC)$	σ^2
Total	IJKn-1	$s^2 = \sum_i \sum_j \sum_k \sum_\ell x^2_{ijk\ell} - C.F.$	

where $C.F. = (IJKn)^{-1} \sum_i \sum_j \sum_k \sum_\ell x_{ijk\ell}$

In actual problems it is more realistic to assume that the $\alpha^B_{j(i)}$ are independently normally distributed with means 0 and the same variance $\sigma^2_{B(A)}$, the $\alpha^{BC}_{jk(i)}$ are independently normally distributed with means 0 and the same variance $\sigma^2_{BC(A)}$ and the $\alpha^B_{j(i)}$, the $\alpha^{BC}_{jk(i)}$, the e_{ijke} are independent. Such assumptions in model (6.53) reduces it to one what is called the mixed-effect model. Under these assumptions the expected values of different mean sum of squares are given in Table 6.4. We refer to Scheffe (1959), Section 8.2 for details. The notations used here are

$$\sigma_A^2 = (I-1)^{-1} \sum_i (\alpha_i^A)^2, \qquad \sigma_{B(A)}^2 = [I(J-1)]^{-1} \sum_i \sum_j (\alpha_{j(i)}^B)^2,$$

$$\sigma_C^2 = (K-1)^{-1} \sum_k (\alpha_k^C)^2, \qquad \sigma_{BC(A)}^2 = [I(J-1)(K-1)]^{-1} \sum_i \sum_j \sum_k (\alpha_{jk(i)}^{BC})^2 .$$

6.8. SPLIT-PLOT DESIGNS

In agricultural experimentation sometimes it is necessary to divide a piece of land into several strips and then apply to each of these strips a different treatment. For convenience we shall call these strips as main plots and the treatments applied to them as main treatments or levels of the main factor. The main plots are subdivided into subplots and one subtreatment is applied to each subplot. In a split plot designs the main effect of the main factor are completely confounded with main plot differences and it is a special case of the crossed and nested classification.

Example 6.5. The following experiment was conducted in Sagar-Bishnupur, West Bengal, India, to study the combined effect on rice crop of nitrogeneous fertilizers at four different levels (taken from two different sources each at two levels) N_1 (source I containing 30 lb. of nitrogen per acre), N_2(source 1 containing 60 lbs. of nitrogen per acre), N_3 (source 2 containing 30 lbs. of nitrogen per acre), N_4 (source 2 containing 60 lbs. of nitrogen per acre); and phosphorus and potash mixtures at 4 different levels

P_1: 0 lb of phosphorus in the form P_2O_5 and 9 lb of potash in the form K_2O per acre,

P_2: 0 lb of phosphorus in the form P_2O_5 and 90 lb of potash in the form of K_2O per acre,

P_3: 0 lb of phosphorus in the form P_2O_5 and 30 lb of potash in the form K_2O per acre,

P_4: 30 lb of phosphorus in the form P_2O_5 and 30 lb of potash in the form K_2O per acre.

The layout adopted was a split plot design with 4 replications. The actual layout and yield of paddy in kilogram per plot is given in Table 6.5.

Table 6.5. Layout and yield of rice crop

Replication 1					Replication 2				
Main treatment		Subtreatments			Main treatment		Subtreatments		
N_2	P_2 6.3	P_1 4.8	P_3 5.0	P_4 5.1	N_4	P_4 5.6	P_1 5.0	P_3 5.4	P_2 4.6
N_3	P_2 5.0	P_4 4.2	P_3 4.2	P_1 4.9	N_1	P_2 5.2	P_4 5.1	P_1 4.3	P_3 5.6
N_1	P_2 4.3	P_1 5.2	P_4 4.6	P_3 5.3	N_3	P_1 5.3	P_2 5.0	P_4 5.1	P_3 4.8
N_4	P_4 5.7	P_2 4.1	P_3 4.9	P_1 5.5	N_2	P_4 5.8	P_2 4.5	P_3 4.6	P_1 5.1

Replication 3					Replication 4				
Main treatments		Subtreatments			Main treatments		Subtreatments		
N_4	P_4 5.3	P_2 4.8	P_1 5.2	P_3 6.2	N_2	P_4 5.3	P_3 4.2	P_1 4.0	P_2 5.1
N_2	P_2 4.9	P_3 5.2	P_1 3.7	P_4 4.7	N_4	P_2 4.9	P_4 4.5	P_3 3.8	P_1 5.1
N_3	P_1 5.3	P_2 5.7	P_3 5.6	P_4 5.3	N_1	P_4 3.8	P_1 3.2	P_2 3.7	P_3 4.2
N_1	P_2 4.7	P_3 4.8	P_1 5.9	P_4 4.3	N_3	P_3 5.7	P_4 5.4	P_1 5.5	P_2 5.1

Calculation of sums of squares of the main effect of nitrogen and the main plot error. This is usually done by completing the analysis of the following two-way classification with nitrogeneous fertilizer and replication as factors (4 observation per cell). The number in each cell indicates the total of 4 observations in the corresponding cell.

	R_1	R_2	R_3	R_4
N_1	19.4	20.2	19.7	14.9
N_2	21.2	20.0	18.5	18.6
N_3	18.3	20.2	21.9	21.7
N_4	20.2	20.6	21.5	18.3

Calculation of sums of squares of the main effect of P and the interaction NP. This is usually done by completing the analysis of a two-way classification with nitrogenous fertilizer and P fertilizer as factors. The number in each cell gives the total of 4 observations in the corresponding cell.

	P_1	P_2	P_3	P_4
N_1	18.6	17.9	19.9	17.8
N_2	17.6	20.8	19.0	20.9
N_3	21.0	20.8	20.3	20.0
N_4	20.8	18.4	20.3	21.1

Table 6.6. Analysis of variance of the split plot experiment

Sources of variation	d.f.	sum of squares	F
Replications	3	2.55	-
Nitrogenous fertilizer	3	2.22	1.01
Main plot error	9	6.65	-

Table 6.6 (contd.)

Between main plots	15	11.42	-
P fertilizers	3	0.18	0.22
Interaction(NP)	10	3.65	1.48
Sub plot error	35	9.81	-
Total	63	25.06	-

6.9. LATTICE DESIGNS

Lattice designs were first introduced by Yates (1937a) and have been investigated extensively by Kempthorne, Federer and their associates. We refer to Kempthorne (1952) and Federer (1955) for an exhaustive account of these designs. In general lattice designs are incomplete block designs where treatments are not related in a special way as are a factorial set of treatments. But in specific cases they can also be treated as incomplete factorial experiments.

The lattice designs are two associate PBIB designs with L_i association scheme. In lattice designs the number of treatments ν is a perfect square. Let $\nu = n^2$. We arrange the n^2 treatments in an $n \times n$ array. In the first replicate two treatments are in the jth block if and only if they are in the jth column of the array. In the second replicate the rows form the blocks. For a third replicate a latin square of order n is superimposed on the array and two treatments are in the same block (of the 3rd replicate) if and only if they correspond to the same letter of the latin square. For further replications we use different orthogonal latin squares of order n.

A design with two replicates, the first one corresponding to the columns as blocks and the second one corresponding to the rows of the array as blocks is called a simple lattice design. Thus a simple lattice design is a two associate PBIB design with L_2 association scheme. A two associate PBIB design with L_3 association scheme is called a triple lattice design. A triple lattice design has three replicates, the columns of the array form the blocks of the first replicate, the rows of the array form the blocks of the second replicate and the letters of the superimposed latin square form the blocks of the third replicate. Simple lattice design is a PBIB design with parameters $\nu = n^2$, $b = 2n$, $r = 2$, $k = n$, $\lambda_1 = 1$, $\lambda_2 = 0$,

$$P^1 = (p^1_{ij}) = \begin{pmatrix} n-2, & n-1 \\ n-1, & (n-1)(n-2) \end{pmatrix}, \quad P^2 = (p^2_{ij}) = \begin{pmatrix} 2 & , 2(n-2) \\ 2(n-2), & (n-1)^2+1 \end{pmatrix}$$

A triple lattice design is a PBIB design with parameters $\nu = n^2$, $b = 3n$, $r = 3$, $k = n$, $\lambda_1 = 1$, $\lambda_2 = 0$,

$$P^1 = \begin{pmatrix} n, & 2(n-2) \\ 2(n-2), & (n-2)(n-3) \end{pmatrix}, \quad P^2 = \begin{pmatrix} 6 & , 3(n-3) \\ 3(n-3), & (n-3)^2+1 \end{pmatrix}.$$

Example 6.6. Let $n = 3$. Consider the array

$$\begin{matrix} 1, & 2, & 3 \\ 4, & 5, & 6 \\ 7, & 8, & 9 \end{matrix}.$$

The blocks of the first replicate are (1,4,7), (2,5,8), (3,6,9). The blocks of the second replicate are (1,2,3), (4,5,6), (7,8,9). These two replicates form a simple lattice design. Superimpose the latin square

$$\begin{matrix} \alpha, & \beta, & \gamma \\ \beta, & \gamma, & \alpha \\ \gamma, & \alpha, & \beta. \end{matrix}$$

on the array. The blocks of the third replicate of the triple lattice design are (1,6,8), (2,4,9), (3,5,7).

If the basic L_i lattice design (i.e. lattice design with L_i association scheme) is replicated s times then we get an incomplete block design with parameters $\nu = n^2$, $b = nsi$, $r = si$, $k = n$, $\lambda_1 = s$, $\lambda_2 = 0$ which can be easily shown to be a two associate PBIB design with L_i association scheme.

In a cubic (three dimensional) lattice design $\nu = n^3$ treatments are arranged into 3 replicates of n^3 treatments in incomplete blocks of size n. The construction and analysis of these designs are fully discussed by Federer (1955) (see also Yates (1939)).

It can be established that a cubic lattice design with $\nu = n^3$ is a three-associate PBIB design with parameters $\nu = n^3$, $b = 3n^2$, $r = 3$, $k = n$, $n_1 = 3(n-1)$, $n_2 = 3(n-1)^2$, $n_3 = (n-1)^3$, $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 0$,

$$P^1 = \begin{pmatrix} n-2 & ,2(n-1) & ,0 \\ 2(n-1, & 2(n-1)(n-2), & (n-1)^2 \\ 0 & ,(n-1)^2 & ,(n-1)^2(n-2) \end{pmatrix},$$

$$P^2 = \begin{pmatrix} 2 & ,2(n-2) & ,(n-1) \\ 2(n-2), & 2(n-1)+(n-2)^2, & 2(n-1)(n-2) \\ (n-1) & ,2(n-1)(n-2) & ,(n-1)(n-2)^2 \end{pmatrix},$$

$$P^3 = \begin{pmatrix} 0 & ,3 & ,3(n-2) \\ 3 & ,6(n-2) & ,3(n-2)^2 \\ 3(n-2), & 3(n-2)^2, & (n-2)^3 \end{pmatrix}.$$

6.9.1. Rectangular Lattice Designs: The rectangular lattice designs were first proposed by Yates (1937a, 1937b). In these designs the number of treatments is of the form $\nu = sn$ and has incomplete blocks of size s in one replicate and has incomplete blocks of size n in the second replicate. The analysis and discussion of rectangular designs for $\nu = p(p-1)$ treatments in incomplete blocks of size p-1 in all replications are given by Harshbarger (1947, 1949, 1951). A rectangular lattice design with $\nu = p(p-1)$, where $p(p-1)$ symbols are identified by a pair (i,j), $i \neq j$, $i,j = 1,\ldots,p$ and all symbols having the same value for i form the i-th block of the first replicate and all symbols having the same value for j form the j-th block of the second replicate, is called a simple rectangular lattice design. Nair (1951) has shown that such simple rectangular lattice designs form a four-associate PBIB designs with parameters $\nu = p(p-1)$, $b = 2p$,

$k = p-1,\ r = 2,\ \lambda_1 = 1,\ \lambda_2 = 0,\ \lambda_3 = 0,\ \lambda_4 = 0,\ n_1 = 2(p-2),$
$n_2 = (p-2)(p-3),\ n_3 = 2(p-2),\ n_4 = 1,$

$$p1 = \begin{pmatrix} p-3, & p-3 & ,1 & ,0 \\ p-3, & (p-3)(p-4), & p-3, & 0 \\ 1 & ,p-3 & ,p-3, & 1 \\ 0 & ,0 & ,1 & ,0 \end{pmatrix},\ p^2 = \begin{pmatrix} 2 & ,2(p-4) & ,2 & ,0 \\ 2(p-4), & (p-4)(p-5), & 2(p-4), & 1 \\ 2 & ,2(p-4) & ,2 & ,0 \\ 0 & ,1 & ,0 & ,0 \end{pmatrix}$$

$$p^3 = \begin{pmatrix} 1 & ,p-3 & ,p-3, & 1 \\ p-3, & (p-3)(p-4), & p-3, & 1 \\ p-3, & p-3 & ,1 & ,0 \\ 1 & ,0 & ,0 & ,0 \end{pmatrix},\ p^4 = \begin{pmatrix} 0 & ,0 & ,2(p-2), & 0 \\ 0 & ,(p-2)(p-3) & ,0 & ,0 \\ 2(p-2), & 0 & ,0 & ,0 \\ 0 & ,0 & ,0 & ,0 \end{pmatrix}$$

In a triple rectangular lattice design $\nu = p(p-1)$ and each treatment is identified by ordered triplets (i,j,k), $i \neq j \neq k$, $i,j,k = 1,\ldots,p$ such that all treatments having the same i value form the ith block of the first replicate, all treatments having the same j value form the jth block of the second replicate and all treatments having the same kth value form the kth block of the third replicate. Nair (1951) has shown that a triple lattice design is a two associate PBIB designs for $p = 3$, a three-associate PBIB design for $p = 4$. For $p \geq 3$ they are not PBIB designs.

The analysis with and without recovery of inter-block information of lattice designs discussed above (except the case of triple rectangular lattice designs) can be treated in the same way as the PBIB designs.

For the analysis of covariance, missing values and efficiency of designs discussed in this chapter and for some other relevant interesting designs we refer to Federer (1955), Kempthorne (1951), Johns (1971) and Raghava Rao (1971).

EXERCISE 6

1. With a view to study the effect of application of nitrogen, phosphorus and potash independently and in combination, a factorial experiment with two levels of nitrogen in the form of ammonium sulphate, two levels of phosphorus in the form of superphosphate and two levels of potash was conducted in a randomized block layout with four replications on paddy crop in Tenali, Andhra Pradesh, India. Details of the experiment are given below.

Two levels of nitrogen: $n_0 = 0$, $n_1 = 44.83$ kilogram per acre.
Two levels of phosphorus: $P_0 = 0$, $p_1 = 67.25$ kilogram per acre.
Two levels of potash: $k_0 = 0$, $k_1 = 67.25$ kilogram per acre.
Plot size 10.4 × 9.0 square meter.

Yield of paddy in kilogram per plot

Treatments	Replications			
	1	2	3	4
$n_0p_0k_0$	3.905	2.035	1.342	0.642
$n_0p_1k_1$	2.005	0.562	0.345	1.570
$n_0p_1k_0$	0.941	2.155	0.930	0.480
$n_0p_0k_1$	2.492	0.533	0.725	2.437
$n_1p_0k_0$	2.651	1.007	1.961	0.289
$n_1p_1k_0$	1.902	2.032	0.767	2.202
$n_1p_0k_1$	1.622	0.285	1.705	2.232
$n_1p_1k_1$	1.206	1.308	1.761	0.267

Analyse the data and interpret the result.

2. Table 6.7 gives the plan and yield of jute fibre inkg per plot of a 2^4 field experiment on jute conducted during 1966 in West Bengal, India. The factors were Cow dung D with levels $n_0 = 0$, $d_0 = 0$, $d_1 = 10$ tons per acre, Nitrochalk N with levels $n_0 = 0$, phosphate P in the form of P_2O_5 with two $n_1 = 0.4$ cwt per acre, phosphate P in the form of P_2O_5 with two levels $P_0 = 0$, $P_1 = 0.6$ cwt per acre and potash K with two levels $k_0 = 0$, $k = 1.0$ cwt per acre.

Table 6.7. Plan of the 2^4 layout and yield of jute fibre per plot

Replication 1

Block 1

$n_0p_1k_0d_0$	$n_0p_1k_1d_0$	$n_0p_0k_0d_1$	$n_1p_1k_1d_0$
4.5	5.5	5.3	3.6
$n_1p_1k_0d_1$	$n_1p_0k_1d_1$	$n_0p_1k_1d_1$	$n_1p_0k_0d_0$
4.1	4.8	5.5	4.2

Block 2

$n_0p_1k_0d_1$	$n_1p_0k_1d_0$	$n_0p_0k_1d_1$	$n_0p_1k_1d_0$
5.0	4.4	4.3	5.1
$n_1p_1k_1d_1$	$n_0p_0k_0d_0$	$n_1p_0k_0d_1$	$n_1p_1k_0d_0$
4.4	5.7	4.0	4.9

Replication 2

Block 1

$n_1p_1k_1d_0$	$n_0p_0k_0d_1$	$n_0p_1k_0d_0$	$n_1p_0k_1d_1$
4.2	4.1	3.8	3.3
$n_1p_0k_0d_0$	$n_1p_1k_0d_1$	$n_0p_0k_1d_0$	$n_0p_1k_1d_1$
4.6	5.1	4.9	4.4

Block 2

$n_1p_0k_1d_0$	$n_0p_1k_0d_1$	$n_0p_0k_0d_0$	$n_1p_1k_0d_0$
4.3	5.2	5.6	3.9
$n_0p_1k_1d_0$	$n_0p_0k_1d_1$	$n_1p_1k_1d_1$	$n_1p_0k_0d_1$
5.5	5.2	5.4	4.2

(a) Identify the confounded interaction.

(b) Analyse the data and interpret the results.

3. Table 6.8 gives the yields of paddy in kilogram per plot in a simple lattice experiment with 25 varieties. The variety number is given within brackets.

Table 6.8. Yield of paddy in two replication with simple lattice design

Replication 1

Block	Yields				
1	(11) 50.8	(12) 88.2	(13) 50.3	(14) 73.9	(15) 49.2
2	(21) 66.5	(22) 41.6	(23) 76.6	(24) 45.3	(25) 83.8
3	(31) 69.4	(32) 71.0	(33) 59.1	(34) 39.7	(35) 64.2
4	(41) 82.8	(42) 50.6	(43) 67.3	(44) 51.3	(45) 82.1
5	(51) 68.1	(52) 72.3	(53) 65.7	(54) 46.7	(55) 79.7

Replication 2

Block	Yields				
1	(11) 55.4	(21) 88.4	(31) 75.4	(41) 94.8	(51) 52.1
2	(12) 67.7	(22) 50.6	(32) 68.1	(42) 48.5	(52) 70.5
3	(13) 60.9	(23) 66.8	(33) 59.5	(43) 54.3	(53) 73.7
4	(14) 58.1	(24) 60.0	(34) 49.9	(44) 79.8	(54) 47.4
5	(15) 49.6	(25) 62.7	(35) 72.7	(45) 52.7	(55) 69.8

Analyze the data to find the high yielding varieties of paddy.

REFERENCES

Das, M.N. (1964). A somewhat alternative approach for the construction of symmetrical factorial designs and obtaining maximum number of factors, Calcutta Stat. Assoc. Bull., 13, pp.1-17.

Das, M.N. and Rao, P.S. (1967). Construction and analysis of some new series of confounded asymetrical factorial designs, Biometrics, 23, pp. 813-822.

Federer, W.T. (1955). Experimental design. The Macmillan Co., N.Y.

Harshbarger, B. (1947). Rectangular lattices, Va. Agri. Exp. Sta. Mem. 1, pp. 1-26.

Harshbarger, B. (1949). Triple rectangular lattices, Biometrics, 5, pp. 1-13.

Harshbarger, B. (1951). Near balanced rectangular lattices, Va. J. Sci. New series, 2, pp.13-17.

Kempthorne, O. (1951). Design and analysis of experiments, Wiley, N.Y.

Lehmann, E. (1959). Testing statistical hypotheses, Wiley, N.Y.

Nair, K.R. (1951). Rectangular lattices and partially balanced incomplete block designs, Biometrics, 7, 145-154.

Raghava Rao, D. (1971). Constructions and combinatorial problems in design of experiments, Wiley, N.Y.

Scheffe, H. (1959). The analysis of variance, Wiley, N.Y.

Yates, F. (1934). The principles of orthogonality and confounding in replicated experiments, J. Agri. Sci., 23, pp. 108-145.

Yates, F. (1936). Complex experiments, J. Roy. Stat. Soc. Suppl. 2, pp. 181-247.

Yates, F. (1937). The design and analysis of factorial experiments, Imp. Bur. soil.-sci. Tech.Comm., 35, pp.1-95.

Yates, F. (1937a). A further note on the arrangement of variety trial: Quasi latin squares, Ann. Eugenics, 7, pp.319-322.

Yates, F. (1939). The recovery of inter-block informations in variety trials arranged in 3-dimensional lattice, Ann.Eugenics 9, pp.136-156.

CHAPTER 7. MULTIVARIATE GENERAL LINEAR HYPOTHESIS WITH APPLICATIONS

In this chapter we shall deal with multivariate general linear hypothesis of the univariate general linear hypothesis of Chapter 3 to vector variates. Unlike the univariate case there is more latitude in the choice of test criterion in the multivariate case although the distribution of different test criteria are quite involved. As we have done in the case of univariate general linear hypothesis we shall treat here two different forms of the multivariate general linear hypothesis and their testing techniques through the multivariate analysis of variance (MANOVA). The first one involves an abstract formulation while the second one involves the multivariate regression model and is suitable for analyzing multivariate design models. As applications we shall treat some multivariate complete block designs.

7.1. GENERAL LINEAR HYPOTHESIS (ABSTRACT FORMULATION)

Let $\underline{X}^{\alpha} = (X_{\alpha 1},\ldots,X_{\alpha p})'$, $\alpha = 1,\ldots,N$ be N independently distributed p-variate normal vectors with means $E(\underline{X}^{\alpha}) = \underline{\mu}^{\alpha} = (\mu_{\alpha 1},\ldots,\mu_{\alpha p})'$ and a common covariance positive definite matrix Σ. Let

$$X = \begin{pmatrix} X_{11},\ldots,X_{1p} \\ X_{21},\ldots,X_{2p} \\ \cdot \quad ,\ldots, \quad \cdot \\ X_{N1},\ldots,X_{Np} \end{pmatrix} = (\underline{X}^1,\ldots,\underline{X}^N)',\ \mu = \begin{pmatrix} \mu_{11},\ldots,\mu_{1p} \\ \mu_{21},\ldots,\mu_{qp} \\ \cdot \quad ,\ldots, \quad \cdot \\ \mu_{N1},\ldots,\mu_{Np} \end{pmatrix} = (\underline{\mu}^1,\ldots,\underline{\mu}^N)'$$

$$\underline{X}_i = (X_{1i},\ldots,X_{Ni})',\ \underline{\mu}_i = (\mu_{1i},\ldots,\mu_{Ni})',\ i = 1,\ldots,p. \qquad (7.1)$$

A multivariate general linear hypothesis model is defined in terms of two linear subspaces π_{Ω} and π_{ω} of dimensions s and s-r respectively, such that $0 \leq s-r < s < N$. It is assumed apriori that $\underline{\mu}_i \in \pi_{\Omega}$ for all i and we are interested to test the null hypothesis H_o that they lie in π_{ω}. We shall assume throughout that $N-s \geq p$ so that we have enough degrees of freedom to estimate Σ.

As in the univariate case (see Chapter 3) this problem can be reduced to a canonical form by applying to each of the N-vector $\underline{X}_i$, i = 1,...,p an orthogonal transformation

$$X \rightarrow Y = OX \tag{7.2}$$

where O is an N × N orthogonal matrix

$$O = \begin{pmatrix} O_{11},\dots,O_{1N} \\ \dots,\dots,\ . \\ O_{N1},\dots,O_{NN} \end{pmatrix} \tag{7.3}$$

such that its first s row vectors $\underline{O}^i = (O_{i1},\dots,O_{iN})$, i =1,...,s span π_Ω with $\underline{O}^{r+1},\dots,\underline{O}^s$ spanning π_Ω. Write

$Y = (\underline{Y}^1,\dots,\underline{Y}^N)'$, $\underline{Y}^\alpha = (Y_{\alpha 1},\dots,Y_{\alpha p})'$, $\alpha = 1,\dots,N$.

Since

$$(\underline{Y}^\alpha)' = (\underline{O}^\alpha \underline{X}_1,\dots,\underline{O}^\alpha \underline{X}_p),$$

and

$$E(\underline{X}_i) = \underline{\mu}_i = \sum_{i=1}^{s} a_i \underline{O}^i$$

where a_i, i = 1,...,s are real constants and all are not equal to zero, $E(\underline{Y}^\alpha) = \underline{0}$ for α= s+1,...,N if and only if all $\underline{\mu}_i \in \pi_\Omega$ and $E(\underline{Y}^\alpha) = \underline{0}$ for α = 1,...,r, s+1,...,N if and only if $\underline{\mu}_i \in \pi_\omega$. Furthermore since

$$Y_{\alpha i} = \sum_{\gamma=1}^{N} O_{\alpha\gamma} X_{\gamma i}, \qquad Y_{\beta j} = \sum_{\delta=1}^{N} O_{\beta\delta} X_{\delta j}$$

and $\mathrm{Cov}(X_{\gamma i}, X_{\delta j}) = \sigma_{ij}$ when γ = δ and is equal to 0 when γ ≠ δ,

$$\mathrm{Cov}(Y_{\alpha i}, Y_{\beta j}) = \sum_{\gamma=1}^{N} \sum_{\delta=1}^{N} O_{\alpha\gamma} O_{\beta\gamma} \mathrm{Cov}(X_{\gamma i}, X_{\delta j})$$

$$= \sigma_{ij} \sum_{\gamma=1}^{N} O_{\alpha\gamma}\, O_{\beta\gamma} = \begin{cases} \sigma_{ij}, & \text{if } \alpha = \beta, \\ 0, & \text{if } \alpha \neq \beta. \end{cases}$$

Thus the row vector of Y are independently distributed normal p-vectors with the same covariance matrix Σ and under π_Ω

$$E(\underline{Y}^{\alpha}) = \begin{cases} \underline{\nu}^{\alpha}, \text{ (say)}, & \alpha = 1,\ldots,s, \\ \underline{0}, & \alpha = s+1,\ldots,N, \end{cases}$$

and under π_{Ω}

$$E(\underline{Y}^{\alpha}) = \begin{cases} \underline{\nu}^{\alpha}, & \alpha = r+1,\ldots,s, \\ \underline{0}, & \alpha = 1,\ldots,r,\ s+1,\ldots,N. \end{cases}$$

Hence the canonical form of the original problem can be stated as follows: $\underline{Y}^{\alpha}$, $\alpha=1,\ldots,N$ are independently distributed p-vectors with the same covariance matrix Σ (positive definite) and with $E(\underline{Y}^{\alpha}) = \underline{0}$, $\alpha = s+1,\ldots,N$. We are interested in testing the null hypothesis H_0: $\underline{\nu}^{\alpha} = E(\underline{Y}^{\alpha}) = \underline{0}$, $\alpha = 1,\ldots,r$. The likelihood of the observations $\underline{y}^{\alpha}$ on $\underline{Y}^{\alpha}$, $\alpha = 1,\ldots,N$, which for given $\underline{y}^{\alpha}$ depends only in $\underline{\nu}^{1},\ldots,\underline{\nu}^{s}$, Σ is

$$L(\underline{\nu}^{1},\ldots,\underline{\nu}^{s}, \Sigma) = (2\pi)^{-Np/2}(\det \Sigma)^{-N/2} \times$$

$$\exp\{-\tfrac{1}{2}\mathrm{tr}\,[\sum_{\alpha=1}^{s}(\underline{y}^{\alpha}-\underline{\nu}^{\alpha})(\underline{y}^{\alpha}-\underline{\nu}^{\alpha})' + \sum_{\alpha=s+1}^{N}\underline{y}^{\alpha}\underline{y}^{\alpha'}]\} \qquad (7.4)$$

Before we proceed to find the likelihood ratio test of H_0 we need to prove the following lemma which will be useful in the sequel.

Lemma 7.1. Let A be a symmetric p × p positive definite matrix and let

$$f(A) = c\ (\det A)^{N/2} \exp\{-\tfrac{1}{2}\,\mathrm{tr}\ A\} \qquad (7.5)$$

where c is a constant. Then f(A) has a maximum in the space of all positive definite matrices when A = NI, where I is the p × p identity matrix.

Proof. Clearly

$$f(A) = c \prod_{i=1}^{p} \theta_i^{N/2} \exp\{-\tfrac{1}{2}\theta_i\}$$

where $\theta_1,\ldots,\theta_p$ are the characteristic roots of the matrix A. Differentiating f(A) with respect to θ_i it is easy to conclude that f(A) is maximum when $\theta_1=\ldots=\theta_N = N$ which holds if and only if A = NI. Hence f(A) is maximum if A = NI. Q.E.D.

Using Lemma 7.1 we obtain

$$\max_{\pi_\Omega} L(\underline{\nu}^1,\dots,\underline{\nu}^s,\Sigma) = \max_{\Sigma} (2\pi)^{-Np/2}(\det \Sigma)^{-N/2} \times \exp\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}(\sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}) = (\tfrac{2\pi}{N})^{-Np/2}(\det \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime})^{-N/2} \times \exp\{-NP/2\}. \tag{7.6}$$

Under the hypothesis H_o the likelihood function $L(\underline{\nu}^{\alpha},\dots,\underline{\nu}^{s}, \Sigma)$ reduces to

$$L(\underline{\nu}^{r+1},\dots,\underline{\nu}^{s},\Sigma) = (2\pi)^{-Np/2}(\det\Sigma)^{-N/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}[\sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime} + \sum_{\alpha=r+1}^{s} (\underline{Y}^{\alpha}-\underline{\nu}^{\alpha})(\underline{Y}^{\alpha}-\underline{\nu}^{\alpha})' + \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}] \tag{7.7}$$

From (7.7) using Lemma 7.1 we obtain

$$\max_{\pi_\omega} L(\underline{\nu}^{r+1},\dots,\underline{\nu}^{s},\Sigma) = \max_{\Sigma}(2\pi)^{-Np/2}(\det\Sigma)^{-N/2} \times \exp\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}[\sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime} + \sum_{\alpha=1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}]\} = (\tfrac{2\pi}{N})^{-Np/2} \times (\det(\sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime} + \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}))^{-N/2} \exp\{-N/2\}. \tag{7.8}$$

From (7.6) and (7.8) the likelihood ratio test of H_o: $\underline{\nu}_1=\dots=\underline{\nu}_r = \underline{0}$ rejects H_o whenever

$$\lambda = (\frac{\det b}{\det(a+b)})^{N/2} \le c' \quad \text{or equivalently} \quad u = \frac{\det b}{\det(a+b)} \le c \tag{7.9}$$

where the constants c, c' are chosen in such a way that the corresponding test has size α and $a = \sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}$, $b = \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}$. This result is due to Wilks (1932) and Hsu (1941) and is popularly called the Wilks' criterion. From Theorem 2.12 and Equation (2.25) $A = \sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}$, $B = \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}$ are independently distributed Wishart matrices provided that $N-s \ge p$ and $r \ge p$ and B has a central Wishart distribution with parameter Σ and N-s degrees of freedom. Under H_o A has central Wishart distribution with parameter Σ and r degrees of freedom whereas under H_1 it is distributed as a noncentral Wishart distribution.

In applications to specific problems it is not always straightforward to carry out the reduction to canonical form, just described, explicitly. However the test statistic U (with values u in (7.9)) can be expressed in terms of the original random variable X by the following argument:

Let $\underline{\hat{\mu}}_i = (\hat{\mu}_{i1},\ldots,\hat{\mu}_{iN})'$ and $\underline{\hat{\hat{\mu}}}_i = (\hat{\hat{\mu}}_{i1},\ldots,\hat{\hat{\mu}}_{iN})'$ be the projections of the vector $\underline{X}_i = (X_{i1},\ldots,X_{iN})'$, i = 1,...,p on π_Ω and π_ω respectively. For each i

$$\sum_{\alpha=1}^{N} (X_{\alpha i} - \hat{\mu}_{\alpha i})^2 = (\underline{X}_i - \underline{\hat{\mu}}_i)' \, (\underline{X}_i - \underline{\hat{\mu}}_i)$$

is the inner product of two vectors, each of which is the difference of the given vector $\underline{X}_i$ and is projection $\underline{\hat{\mu}}_i$ on π_Ω and it remains unchanged under the orthogonal transformation of the coordinate system in which the variables are expressed. Now $O\underline{X}_i$ can be interpreted as expressing $\underline{X}_i$ in a new coordinate system with the first s coordinate axes lying in π_Ω. Hence the projection of the transformed vector $\underline{Y}_i$ on π_Ω is $(Y_{1i},\ldots,Y_{si},0,\ldots,0)$ so that the difference between the vector $\underline{Y}_i$ and its projection on π_Ω is $(0,\ldots,0,Y_{s+1,i},\ldots,Y_{Ni})$. Therefore the (i,j)th element of $\sum_{\alpha=s+1}^{N} \underline{Y}^\alpha \underline{Y}^{\alpha'}$ is given by

$$\sum_{\alpha=s+1}^{N} Y_{\alpha i} Y_{\alpha j} = \sum_{\alpha=1}^{N} (X_{\alpha i} - \hat{\mu}_{\alpha i})(X_{\alpha j} - \hat{\mu}_{\alpha j}). \tag{7.10}$$

In the same way the difference between the projections of $\underline{Y}_i$ on π_Ω and π_ω is given by $(Y_{1i},\ldots,Y_{ri},0,\ldots,0)$. Thus $\sum_{\alpha=1}^{r} Y_{\alpha i} Y_{\alpha j}$ is equal to the inner product of the difference of these projections. In other words

$$\sum_{\alpha=1}^{r} Y_{\alpha i} Y_{\alpha j} = (\underline{\hat{\mu}}_i - \underline{\hat{\mu}}_i)' \, (\underline{\hat{\mu}}_i - \underline{\hat{\hat{\mu}}}_i). \tag{7.11}$$

7.1.1. Invariance of Multivariate General Linear Hypothesis:

Let $\Lambda = (\underline{\nu}^1,\ldots,\underline{\nu}^r)$. In terms of the transformed vector $\underline{Y}^\alpha$,

$\alpha=1,\dots,N$, the problem of testing H_o: $\Lambda = 0$ against H_1: $\Lambda \neq 0$ remain invariant under the following three groups of transformations:

(1) The group T of translations which translates $\underline{Y}^{\alpha} \to \underline{Y}^{\alpha} + \underline{d}^{\alpha}$, $\alpha=r+1,\dots,s$, $\underline{d}^{\alpha} = (d_{\alpha 1},\dots,d_{\alpha p})' \in T$. The maximal invariant in the space of Y under T is $(\underline{Y}^1,\dots,\underline{Y}^r, \underline{Y}^{s+1},\dots,\underline{Y}^N)$.

(2) Let Z be a $r \times p$ matrix such that $Z' = (\underline{Y}^1,\dots,\underline{Y}^r)$ and let W be the $(N-s) \times p$ matrix such that $W' = (\underline{Y}^{s+1},\dots,\underline{Y}^N)$. The group $O(r)$ of $r \times r$ orthogonal matrices, operating as $Z \to OZ$, $O \in O(r)$, in the space of Z and the group $O(N-s)$ of $(N-s) \times (N-s)$ orthogonal matrices, operating as $W \to OW$, $O \in O(N-s)$, in the space of W affect neither the independence nor the covariance between any two row vectors of Z and W.

The following lemma will establish that $Z'Z = \sum_{\alpha=1}^{r} \underline{Y}^{\alpha}\underline{Y}^{\alpha}$ is a maximal invariant in the space of Z.

Lemma 7.2. $Z'Z$ is a maximal invariant in the space of Z under $O(r)$.

Proof: Since for $O \in O(r)$, $(OZ)'(OZ)=Z'Z$ the matrix $Z'Z$ is a maximal Z,Z^* invariant if we can establish that for any two elements Z,Z^* in the same space $Z^{*\prime}Z^*=Z'Z$ implies that there exists a matrix $O \in O(r)$ such that $Z^*=OZ$. Let us first consider $r = p$. Without any loss of generality we can assume that the p columns of Z are linearly independent (the exceptional set of Z's for which this does not hold has probability measure 0). Now $Z^{*\prime}Z^* = Z'Z$ implies that $O = Z^*Z^{-1}$ is an orthogonal matrix and that $Z^* = OZ$.

Now consider the case $r > p$. Without any loss of generality we can assume here that the p columns of Z are linearly independent. Since for any two p-dimensional subspaces of the r-space there exists an orthogonal transformation, transforming one to other we assume that after a suitable orthogonal transformation the p column vectors of Z and Z^* lie in the same subspace and the problem is reduced to the case $r = p$.

Lastly consider the case $r < p$. The first r column vectors of Z can be assumed to be linearly independent. Write $Z = (Z_1, Z_2)$ where Z_1, Z_2 are subspaces of dimensions $r \times r$ and $r \times (p-r)$ respectively and similarly for Z^*. Since $Z^{*\prime}Z^* = Z'Z$ we obtain $Z_1^{*\prime}Z_1^* = Z_1'Z_1$, $Z_1^{*\prime}Z_2 = Z_1'Z_2$, $Z_2^{*\prime}Z_2^* = Z_2'Z_2$. Now by the previous argument for the case $r = p$ $Z_1^{*\prime}\ Z_1^* = Z_1'Z_1$ implies that there exists an orthogonal matrix $B = (Z_1^{*\prime})^{-1}\ Z_1'$ such that $Z_1^*=BZ_1$. Also $Z_1^{*\prime}\ Z_2^* = Z_1'Z_2$ implies that $Z_2^* = BZ_2$. Obviously $Z_2^{*\prime}Z_2^*=Z_2'Z_2$ with $Z_2^*=BZ_2$ as B is an orthogonal matrix. Q.E.D.

Similarly a maximal invariant in the space of W under $O(N-s)$ is $W'W = \sum_{\alpha=s+1}^{N} \underline{Y}^{\alpha}\underline{Y}^{\alpha\prime}$. From Theorem 2.12, $W'W$ is positive definite with probability one since $N-s \geq p$ by assumption.

(3) The problem of testing H_0 against H_1 remains invariant under the full linear group $G_\ell(p)$ (the multiplicative group of $p \times p$ nonsingular matrices) of transformations g transforming Z to gZ and W to gW. The corresponding induced transformation in the space of (A,B) is given by $(A,B) \rightarrow (gAg', gBg')$.

Lemma 7.3. The roots of $\det(A-\lambda B) = 0$ (the characteristic roots of AB^{-1}) are maximal invariant in the space of (A,B) under the group $G_\ell(p)$ of transformations g, transforming $(A,B) \rightarrow (gAg', gBg')$.

Proof. Since for $g \in G_\ell(p)$ $\det(g(A-\lambda B)g') = (\det g)^2 \det(A-\lambda B)$ $\det(gAg'-\lambda gBg') = 0$ and $\det(A-\lambda B) = 0$ have the same roots. Thus the roots of $\det(A-\lambda B) = 0$ are invariant under $G_\ell(p)$. To show that they are maximal invariant let us suppose that (A,B), (A^*,B^*) be two elements in the space (A,B) and let $\det(A-\lambda B) = 0$, $\det(A^*-\lambda B) = 0$ have the same roots. We would like to show that there exists a $g \in G_\ell(p)$ such that $A^* = gAg'$, $B^* = gBg'$. Since B, B^* are positive definite with probability one (by the assumption $N-s \geq p$), there exists $g_1, g_2 \in G_\ell(p)$ such that $g_1Bg_1' = I$, $g_2B^*g_2' = I$. Thus

$$\det(g_1Ag_1' - \lambda I) = 0, \quad \det(g_2Ag_2' - \lambda I) = 0$$

have the same roots. Hence we conclude that there exists $p \times p$

orthogonal matrices O_1, O_2 satisfying

$O_1 g_1 A g_1' O_1' \;=\; O_2 g_2 A^* g_2' O_2'$

which implies that $A^* = (g_2^{-1} O_2' O_1 g_1)\, A (g_1' O_1' O_2 g_2'^{-1}) = gAg'$, where $g = g_2^{-1} O_2' O_1 g_1 \in G_\ell(p)$. Furthermore from $O_1 g_1 B g_1' O_1' = O_2 g_2 B^* g_2' O_2'$ we get $B^* = gBg'$.

Q.E.D.

Let $R_1, \ldots, R_p$ be the roots of $\det(A - \lambda B) = 0$. Then $R_1, \ldots, R_p$ are the maximal invariants in the space of (A, B) under $G_\ell(p)$. Similarly a corresponding maximal invariant in the parametric space of $(\underline{\nu}^1, \ldots, \underline{\nu}^s, \Sigma)$ is $\theta_1, \ldots, \theta_p$, the characteristic roots of $\Lambda\Lambda'\Sigma^{-1}$. The statistic U in (7.9) can be written as

$$U = \det(B(A+B)^{-1}) = \prod_{i=1}^{p} (1 + R_i)^{-1} . \tag{7.12}$$

Anderson (1958) called this statistic $U_{p,r,N-s}$. Several invariant tests are proposed for this problem and they are as follows. In all cases below we shall use the constant c as the generic notation of the cut-off point depending on the level of significance α of the test.

(i) Wilks' criterion (Wilks, 1932, Hsu, 1940)

Reject H_o whenever $u = \det(a(a+b)^{-1}) = \prod_{i=1}^{p} (1 + r_i) \leq c.$ (7.13)

(ii) Lawley's v (Lawley, 1938) and Hotelling's T_o^2 (Hotelling, 1951) criterion

$$\text{Reject } H_o \text{ whenever } v = \operatorname{tr} ab^{-1} = \sum_{i=1}^{p} r_i = \frac{T_o^2}{N-p} \geq c. \tag{7.14}$$

(iii) The largest and the smallest root criteria of Roy (Roy, 1957).

$$\text{Reject } H_o \text{ whenever } \max_i r_i \geq c \tag{7.15}$$

$$\text{Reject } H_o \text{ whenever } \min_i r_i \geq c . \tag{7.16}$$

(iv) Pillai's statistic (Pillai, 1955).

$$\text{Reject } H_o \text{ whenever } \operatorname{tr} a(a+b)^{-1} = \sum_{i=1}^{p} (1+r_i)^{-1} \geq c. \tag{7.17}$$

Among all these tests the test (iv) has received much less attention compared to others. It follows from the fact that the distribution of $R_1,\ldots,R_p$ under the alternatives depends continuously on $\theta_1,\ldots,\theta_p$, if $r > 1$ there does not exist a uniformly most powerful test for testing H_o against $H_1 : \Lambda \neq 0$.

7.1.2. Special case $r = 1$. All these tests reduce to Hotelling's T^2-test when $r = 1$. Furthermore they all reduce to F-test when $p = 1$ and the two-tailed Student's t-test when $p = r = 1$. If $r = 1$, then $A = \underline{Y}^1 Y^{1'}$ and $\det(AB^{-1} - \lambda I) = 0$ has only one non-zero root. All coefficients of λ^j, $j < p-1$ in $\det(AB^{-1} - \lambda I) = 0$ therefore vanish and the equation $\det(AB^{-1} - \lambda I) = 0$ becomes $(-\lambda)^p + (-\lambda)^{p-1} \operatorname{tr} AB^{-1} = 0$.

The maximal invariant in the space of (A,B) is

$$R_1 = \operatorname{tr} AB^{-1} = \operatorname{tr} \underline{Y}^1 \underline{Y}^{1'} B^{-1} = \underline{Y}^{1'} B^{-1} \underline{Y}^1 . \tag{7.18}$$

The statistic R_1 is called the Hotelling's T^2-statistic. The distribution of R_1 depends only on θ_1, the maximal invariant in the parametric space and is given by

$$\theta_1 = \operatorname{tr} \underline{\nu}^1 \underline{\nu}^{1'} \Sigma^{-1} = \nu^{1'} \Sigma^{-1} \underline{\nu}^1 = \sigma^2 \text{ (say)} \tag{7.19}$$

From Theorem 2.14 the probability density function of R_1 is given by

$$f_{R_1}(r_1) = \frac{\exp\{-\frac{1}{2}\delta^2\}}{(N-s)\,\Gamma(\frac{1}{2}(N-s-p+1))} \sum_{j=0} \frac{(\delta^2/2)^j \, r_1^{\frac{1}{2}p+j-1}\,\Gamma(\frac{N-s+1}{2} + j)}{j!\,\Gamma(\frac{1}{2}p+j)\,(1+r_1)^{\frac{1}{2}(N-s+1)+j}} . \tag{7.20}$$

Under H_0 $\delta^2 = 0$ and under H_1 $\sigma^2 > 0$.

From Example (2.9) the test (Hotelling's T^2 test) which rejects H_0 whenever $r_1 \geq c$, the constant c depends only on the level of significance α of the test, is uniformly most powerful invariant for testing $H_0 : \sigma^2 = 0$ against the alternatives $\sigma^2 > 0$.

From Example (2.11) we conclude that among all tests of $H_0 : \underline{\nu}^1 = \underline{0}$ against alternatives $\underline{\nu}^1 \neq \underline{0}$ with power functions depending only on σ^2 the Hotelling's T^2 test is uniformly most powerful.

From Stein (1956) or Kiefer and Schwartz (1965) the Hotelling's T^2-test is admissible. From Giri and Kiefer (1965) the Hotelling's T^2-test is locally and asymptotically minimax as $\delta^2 \to 0$ and $\sigma^2 \to \infty$ respectively and most stringent. From Giri, Kiefer and Stein (1963), Linnik, Pliss and Salaevskii (1966), Salaevskii (1968) the Hotelling's T^2-test is minimax (see Section 2.4.4 and original references or Giri 1976, 1977 for details). We shall discuss the optimum properties of all the five test criteria in Section 7.4. Now we shall find the distribution of $R_1,\dots,R_p$ under H_0.

7.2. THE DISTRIBUTION OF $R_1,\dots,R_p$ UNDER H_0

In theory one would be able to derive the null distribution of five test statistic as given in (7.13)-(7.17) from the joint distribution of $R_1,\dots,R_p$ under H_0. Since for any $g \in G_\ell(p)$

$$\det(gAg'-\lambda gBg') = (\det g)^2 \det(A-\lambda B) ,$$

by choosing g such that $g\Sigma g' = I$, we conclude that to find the joint distribution of $R_1,\dots,R_p$ under H_0 we can without any loss of generality assume that $\Sigma = I$. Thus the joint distribution of $R_1,\dots,R_p$ does not depend on Σ under H_0 and depends only on $\theta_1,\dots,\theta_p$, the characteristic roots of $\Lambda\Lambda'\Sigma^{-1}$ under H_1: $\Lambda \neq 0$. From Theorem 2.12, the random matrices A,B are independently distributed as $W_p(\Sigma,r)$ and $W_p(\Sigma,N-s)$ respectively provided $N-s \geq p$, $r \geq p$. Note that the assumptions $N-s \geq p$, $r \geq p$ make both A,B positive definite with probability one. Let $r = n_1$, $N-s = n_2$ and let $R_1 > R_2 > \dots > R_p > 0$ denote the ordered characteristic roots of AB^{-1} (the probability of two roots being equal is 0). Let

$$V_i = \frac{R_i}{1+R_i} , \quad i = 1,\dots, p . \tag{7.21}$$

Obviously $V_1,\dots,V_p$ are the characteristic roots of $A(A+B)^{-1}$, that is, the roots of $\det(A-\lambda(A+B)) = 0$.

Rather than finding the joint distribution of $R_1,\dots,R_p$ directly we will find it convenient to find first the joint distribution of $V_1,\ldots,V_p$ and then derive from it the joint distribution of $R_1,\ldots,R_p$.

Let $C = A+B$ and let V be a diagonal matrix with diagonal elements $V_1,\ldots,V_p$. We can write

$$C = WW', \quad A = WVW'$$

where $W = (W_{ij})$ is a $p \times p$ nonsingular matrix. Obviously the matrix W is not unique. To determine W uniquely we shall require that $W_{i1} > 0$, $i = 1,\ldots,p$ (the probability of $W_{i1} = 0$ is zero). Writing J for the Jacobian, the Jacobian of the transformation $(A,B) \to (W,V)$ is equal to

$$J[(A,B) \to (W,V)] = J[(A,B) \to (A,C)] \times J[(A,C) \to (W,V)] \tag{7.22}$$

It is easy to verify that $J[(A,B) \to (A,C)] = 1$. From Olkin (1952)

$$J[(A,C) \to (W,V)] = 2^p (\det W)^{p+2} \sum_{i<j} (V_i - V_j). \tag{7.23}$$

From (2.25) the joint probability density function of (A,B) under H_0 is (with $\Sigma = I$, $n_1 \geq p$, $n_2 \geq p$)

$$C_{n_1,p}\, C_{n_2,p}\, (\det a)^{(n_1-p-1)/2} (\det b)^{(n_2-p-1)/2} \exp\{-\tfrac{1}{2}\operatorname{tr}(a+b)\} \tag{7.24}$$

where

$$C_{n,p} = \prod_{i=0}^{p-1} \Gamma((n-i-1)/2)\; 2^{(n-1)p/2}\, \pi^{p(p-1)/4}$$

From (7.23)-(7.24) the joint probability density of (W,V) is

$$f_{W,V}(\omega, v) = C_{n_1,p}\, C_{n_2,p} \prod_{i=1}^{p} v_i^{(n_1-p-1)/2} (1-v_i)^{(n_2-p-1)/2} \tag{7.25}$$
$$\times \prod_{i<j} (v_i - v_j) (\det \omega\omega')^{(n_1+n_2-p)/2} \exp\{-\tfrac{1}{2}\operatorname{tr}\,\omega\omega'\}.$$

Integrating out ω in (7.25) the probability density function of V is given by

$$f_V(v) = K C_{n_1,p}\, C_{n_2,p} \prod_{i=1}^{p} [v_i^{(n_1-p-1)/2} (1-v_i)^{(n_2-p-1)/2} \prod_{i<j} (v_i - v_j)] \tag{7.26}$$

where

$$K = (2\pi)^{p^2/2} \int (2\pi)^{-p^2/2}\, 2^p [\det(\omega\omega')]^{(n_1+n_2-p)/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\,\omega\omega'\} d\omega$$

$$= (2\pi)^{p^2/2}\, E[\det WW']^{(n_1+n_2-p)/2} \tag{7.27}$$

and $W = (W_{ij})$, the W_{ij} are independently distributed normal random variables with means 0 and the same variance unity. Thus to evaluate $E[\det WW']^{(n_1+n_2-p)/2}$ in (7.27). We may use the fact that $WW' = S$ (say) is a Wishart matrix $W_p(I,p)$ and by (2.25) its probability density function is given by

$$f_S(s) = C_{p,p}(\det s)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\mathrm{tr}\ s\}\ . \tag{7.28}$$

Note that (7.28) holds only for the evaluation of K as given in (7.27). From (2.25) and (7.27)-(7.28)

$$E(\det WW')^{(n_1+n_2-p)/2}$$

$$= C_{p,p}\int(\det s)^{(n_1+n_2-p-1)/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\ s\}\ ds$$

$$= C_{p,p}/C_{n_1+n_2,p}\ . \tag{7.29}$$

Thus

$$K = \frac{(2\pi)^{p^2/2}\, C_{p,p}}{C_{n_1+n_2,p}}\ . \tag{7.30}$$

Since $dV_i = (1+R_i^2)^{-1}\, dR_i$, from (7.26) the probability density function of R, a diagonal matrix with diagonal elements $R_1,\ldots,R_p$ is

$$f_R(r) = C \prod_{i=1}^{p} r_i^{(n_1-p-1)/2} (1+r_i)^{-(n_1+n_2)/2} \prod_{i<j} (r_i-r_j) \tag{7.31}$$

where

$$c = \frac{\pi^{p/2} \prod_{i=1}^{p} \Gamma(\tfrac12(n_1+n_2-i+1))}{\prod_{i=1}^{p} \Gamma(\tfrac12(n_1-i+1)\ \Gamma(\tfrac12(n_2-i+1))\ \Gamma(\tfrac12(p+1-i))} .$$

7.3. DISTRIBUTION OF TEST STATISTICS UNDER H_0

Since all the four test criteria are functions of R the distribution of all test statistics given in (7.13)-(7.17) under H_0 can be obtained from (7.31). However, the exact evaluation of all these distributions in specific cases are quite involved. Using (7.31) and computing various moments of $U = U_{p,r,N-s}$ under H_0 we obtain

$$E(U^k) = \prod_{i=1}^{p} (X_i^k), \ k = 0,1,\ldots \tag{7.32}$$

where $X_1,\ldots,X_p$ are independently distributed central beta random variables with parameters $(\frac12(N-s-i+1), \frac12 r)$, $i = 1,\ldots,p$. Since U lies between 0 and 1, these moments determine uniquely the distribution of U under H_0. Thus under H_0 U is distributed as $U = \prod_{i=1}^{p} X_i$. Furthermore under H_0 $U_{p,r,N-s}$ and $U_{r,p,N-p-s+r}$ have the same distribution. In the special case $p = 1$, $(N-s)(1-U)/rU$ has central F-distribution with degrees of freedom $(r,N-s)$. For $p = 2$, $(N-s-1)(1-\sqrt{U})/r\sqrt{U}$ has central F-distribution with $(2r, 2(N-s+1))$ degrees of freedom. Schatzoff (1966) has given a method of obtaining the exact distribution of $U_{p,r,N-s}$ using the representation of $-\log_e U_{p,r,N-s}$ as a sum of independent variables and taking successive convolutions and has tabulated the factor for converting χ^2_{pr} percentiles to exact percentiles of $-(N-s-\frac12(p-r+1))\log U_{p,r,N-s}$ for $p = 3,\ldots,8$ and for values of r such that $pr \le 70$ (see Table 7.3). However he has not given the explicit expression of the null distribution of $U_{p,r,N-s}$ while Pillai and Gupta (1969) have given both for $p \le 6$ and have succeeded to remove the restriction $pr \le 70$. Unlike Consul (1966), who following Nair (1938) has used inverse Mellin transform to obtain the distribution of $U_{r,p,N-s}$ for

$p \le 4$ as infinite series, Pillai and Gupta (1969) have obtained the same in finite series form except when both p and r are odd. In the case when p and r are both odd the series became infinite. Pillai and Gupta have also extended Schatzoff's tables for $11 \le r \le 22$ (see Table 7.4). Lee (1972) has studied the probability density function of $U_{p,r,N-s}$ when one of p and r is even with a view to simplifying the expressions obtained by earlier authors for numerical computations. When both p and r are odd he has given expressions in terms of simple integrals and has extended the tables of chisquare correction factors to cover, in conjunction of previous tables, values $p \le r \le 20$ and $pr \le 144$ with omission when one of p and r is odd and is greater than 10. Mathai (1971a,b) and Mathai and Rathie (1971) have also derived the distribution $U_{p,r,N-s}$ using inverse Mellin transform and computed 5% and 1% point of $U_{p,r,N-s}$. For further references connected with the distribution of $U_{p,r,N-s}$ under H_0 we refer to Pillai (1976).

Box (1949) gave an asymptotic expansion for the distribution of a monotone function of the likelihood ratio statistic $\lambda = (U_{p,r,N-s})^{N/2}$ when H_0 is true. The expansion converges extremely rapidly and therefore the level of significance derived from it will be quite adequate for moderate values of N. For large N, the Box result is equivalent to the large sample result of Wilks (1938), that is, under H_0, $-2 \log \lambda$ is distributed as χ^2_{pr} with pr degrees of freedom as $N \to \infty$. The Box approximation (with $p \le r$) is, under H_0,

$$P\{-r \log U_{p,r,N-s} \le z\} = P\{\chi^2_{pr} \le z\}$$

$$+ (\gamma/r^2)\,[P\{\chi^2_{pr+4} \le z\} - P\{\chi^2_{pr} \le z\}] + 0(N^{-4}), \qquad (7.33)$$

where $\gamma = pr(p^2+r^2-5)/48$. If just the first term is used the total error of approximation is $0(N^{-2})$, if both terms are used the error is $0(N^{-4})$. Furthermore if $r < p$ we use the result that under H_0 $U_{p,r,N-s}$ is distributed as $U_{r,p,N-s-p+r}$. Rao (1951,1952) has given another approximation in terms of beta series.

For the Lawley-Hotelling test criterion Pillai (1957,1960) obtained approximate upper 5% and 10% points of $V = \mathrm{tr}\ AB^{-1}$ under H_0 for $p = 2, 3, \ldots, 8$ and for various values of r and N-s. Marrow (1948) has shown that under H_0 $N\ \mathrm{tr}\ AB^{-1}$ is distributed as χ^2_{pr} when $N \to \infty$.

Pillai and Jayachandran (1970) have computed exact percentage points $\mathrm{tr}\ A(A+B)^{-1}$ under H_0 for a few cases corresponding to each of $p = 3$ and $p = 4$.

The largest and the smallest root criteria of Roy can be used in the general case, although percentage points are available only for the restricted values of the parameters. Appropriate tables are given by Foster and Rees (1957), Foster (1957, 1958), Heck (1960) and Pillai (1960). These four test criteria have been compared on the basis of their power functions in some details by Smith et al. (1962) and Gabriel (1969).

7.4. OPTIMUM PROPERTIES OF TESTS OF GENERAL LINEAR HYPOTHESIS

Using the argument that follows Stein's proof of admissibility of Hotelling's T^2 test (Stein 1956) Schwartz (1964a) has shown that for testing $H_0 : \Lambda = 0$ against alternatives $H_1 : \Lambda \neq 0$ the Pillai's test ~~est~~ which rejects H_0 whenever $\mathrm{tr}\ a(a+b)^{-1} \leq c$, where the constant ~~onstant~~ c depends on the level of significance α of the test, is admissible ~~missible~~. He has also obtained the following results:

(i) For testing $H_0 : \Lambda = 0$ against the alternatives $\mathrm{tr}\Lambda\Lambda'\Sigma^{-1} = \delta$ Pillai's test is minimax in the sense of Giri and Kiefer (1964a) as $\delta \to 0$

(ii) Among all level α tests of H_0 which depends only on $R_1, \ldots, R_p$ and which therefore have power functions of the form $\alpha + k\ \mathrm{tr}(\Lambda\Lambda'\Sigma^{-1}) + o(\Lambda\Lambda'\Sigma^{-1})$ Pillai's test minimizes the value of k.

Ghosh (1964), using Stein's approach, has shown that the Lawley-Hotelling's trace test which rejects H_0 whenever $\mathrm{tr}\ ab^{-1} \geq c$ and the Roy's test which rejects H_0 whenever $\max_i r_i \geq c$, where the constant c depends on the level of significance α of the test are admissible for testing $H_0 : \Lambda = 0$ against the alternatives $H_1 : \Lambda \neq 0$. Thus, as a consequence of the following result of Anderson et al. (1964), they are unbiased for this problem.

Anderson et al. (1964) have given sufficient conditions on invariant tests depending only on $R_1, \ldots, R_p$ for the power functions to be monotonically increasing functions of each θ_i, $i = 1, \ldots, p$. Further they have shown that the likelihood ratio test, the Lawley-Hotelling's trace test and the Roy's maximum characteristic roots test satisfy these conditions. The monotonicity of the power function of Roy's test has been demonstrated by Roy and Mikhail (1961) using a geometrical argument.

Kiefer and Schwartz (1965) have shown, using the Bayes approach, that Pillai's test is admissible Bayes for this problem and it is fully invariant, similar and as a consequence of the result of Anderson et al. (1964) given in the preceding paragraph, unbiased. Using the same Bayes approach Kiefer and Schwartz have also proved the admissibility of the likelihood ratio test under the restriction that $N-s \geq p+r-1$, although the admissibility of the likelihood ratio test can be proved without this added restriction (Schwartz, 1964b). Sinha and Giri (1977) proved the Bayes character (and hence admissibility) of the likelihood ratio test whenever $N-s > p$. Narain (1950) has shown that the likelihood ratio test is unbiased. We refer to Nandi (1963) for a related admissibility result and to John (1971) for an optimality result.

7.5. MULTIVARIATE REGRESSION MODEL

We shall discuss here the multivariate regression model of the general linear hypothesis. This formulation is very appropriate for the analysis of multivariate design model. Let $\underline{X}^\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})'$, $\alpha = 1, \ldots, N$ be independently distributed normal p-vectors with

$$E(\underline{X}^\alpha) = \beta \underline{z}^\alpha$$
$$\mathrm{Cov}(\underline{X}^\alpha) = \Sigma, \tag{7.34}$$

lternatives
g result of
problem.
conditions on
e power
of each θ_i,
elihood ratio
y's maximum
. The
as been
metrical

the Bayes
or this
a consequence
he preceding
ch Kiefer
the likelihood
although the
proved without
nd Giri (1977)
y) of the
950) has shown
fer to Nandi
ohn (1971) for

model of the
ry appropriate
t $\underline{x}^\alpha$ =
ributed normal

(7.34)

---

where $\underline{z}^\alpha = (z_{\alpha 1},\dots,z_{\alpha s})'$, $\alpha = 1,\dots,N$ are vectors of known constants, $\beta = (\beta_{ij})$ is a $p \times s$ matrix of unknown elements β_{ij}. We shall assume that $N-s \geq p$ and the rank of the $s \times N$ matrix $z = (z^1,\dots,z^N)$ is s. Let $\beta = (\beta_1, \beta_2)$ where β_1, β_2 are submatrices of dimensions $p \times r$ and $p \times (s-r)$ respectively. We are interested in testing the hypothesis $H_0: \beta_1 = \beta_1^0$, where β_1^0 is a fixed matrix, β_2 and Σ are unknowns. Obviously the dimension of π_Ω is s and that of π_ω is s-r (see Section 7.1 fornotation). The likelihood of the observations $\underline{x}^\alpha$, $\alpha = 1,\dots,N$ is given by

$$L(\beta,\Sigma) = (2\pi)^{-Np/2}(\det \Sigma)^{-N/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}(\sum_{\alpha=1}^{N}(\underline{x}^\alpha-\beta\underline{z}^\alpha)(\underline{x}^\alpha-\beta\underline{z}^\alpha)')\}. \tag{7.35}$$

Let

$$A = zz' = \sum_{\alpha=1}^{N} \underline{z}^\alpha \underline{z}^{\alpha'}, \quad x = (\underline{x}^1,\dots,\underline{x}^N), \quad c = xz',$$
$$X = (\underline{X}^1,\dots,\underline{X}^N). \tag{7.36}$$

Using the results in Section 1.1.10 and (7.35) the maximum likelihood estimate $\hat{\beta}$ of β is given by

$$\hat{\beta}A = c. \tag{7.37}$$

~~Since~~ Since the rank of z is s, A is nonsingular and the unique maximum ~~likelihood~~ likelihood estimate of β is $\hat{\beta} = cA^{-1}$. Thus

$$\max_{\beta,\Sigma} L(\beta,\Sigma) = \max_{\Sigma} (2\pi)^{-Np/2}(\det \Sigma)^{-N/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}(\sum_{\alpha=1}^{N}(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)')\} \tag{7.38}$$

From (7.38), using Lemma 7.1. the maximum likelihood estimate of Σ under π_Ω is

$$\hat{\Sigma} = \frac{1}{N}\sum_{\alpha=1}^{N}(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)' = \frac{1}{N}(\sum_{\alpha=1}^{N}\underline{x}^\alpha\underline{x}^{\alpha'} - \hat{\beta}A\hat{\beta}'). \tag{7.39}$$

Thus

$$\max_{\beta,\Sigma} L(\beta,\Sigma) = (\frac{2\pi}{N})^{-Np/2}(\det(\sum_{\alpha=1}^{N}(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)(\underline{x}^\alpha-\hat{\beta}\underline{z}^\alpha)'))^{-N/2}\exp(-Np/2). \tag{7.40}$$

To find the maximum of the likelihood function under H_0, let

$$z^{\alpha}=\begin{pmatrix} z^{\alpha}_{(1)} \\ z^{\alpha}_{(2)} \end{pmatrix}, \quad z^{\alpha}_{(1)} = (z_{\alpha 1},\dots,z_{\alpha r})', \quad \underline{y}^{\alpha}= \underline{x}^{\alpha} - \beta_1^{o} z^{\alpha}_{(1)},$$

$$\alpha = 1,\dots,N. \tag{7.41}$$

Then under H_o,

$$\underline{\ }^{\alpha} = \underline{x}^{\alpha} - \beta_1^{o} z^{\alpha}_{(1)}, \quad \alpha = 1,\dots,N$$

are independently normally distributed with $E(\underline{y}^{\alpha}) = \beta_2 z_{(2)}$ and the same covariance matrix Σ. Partition

$$\beta = (\hat{\beta}_1,\hat{\beta}_2), \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad C = (C_1,C_2), \quad A = \begin{pmatrix} A_{(11)}, & A_{(12)} \\ A_{(21)}, & A_{(22)} \end{pmatrix}, \tag{7.42}$$

where Z_1 is a $r \times N$ submatrix of Z, C_1 is a $p \times r$ submatrix of C, $A_{(11)}$ is a $r \times r$ submatrix of A. The likelihood function under H_o can be written as

$$L(\beta_2,\Sigma) = (2\pi)^{-Np/2}(\det\Sigma)^{-N/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(\sum_{\alpha=1}^{N}(\underline{y}^{\alpha}-\beta_2 z^{\alpha}_{(2)})$$

$$(\underline{y}^{\alpha}-\beta_2 z^{\alpha}_{(2)})')\}. \tag{7.43}$$

From (7.43) it is thus obvious that the maximum likelihood estimates $\hat{\hat{\beta}}_2$, and $\hat{\hat{\Sigma}}$ of β_2 and Σ respectively under H_o are given by

$$\hat{\hat{\beta}}_2 = \sum_{\alpha=1}^{N} \underline{y}^{\alpha} z^{\alpha}_{(2)} \left(\sum_{\alpha=1}^{N} z^{\alpha}_{(2)} z^{\alpha'}_{(2)}\right)^{-1} = (C_2 - \beta_1^{o} A_{(12)})\, A^{-1}_{(22)}$$

$$\hat{\hat{\Sigma}} = \frac{1}{N}\sum_{\alpha=1}^{N}(\underline{y}^{\alpha}-\hat{\hat{\beta}}_2 z^{\alpha}_{(2)})(\underline{y}^{\alpha}-\hat{\hat{\beta}}_2 z^{\alpha}_{(2)})'$$

$$= \frac{1}{N}\sum_{\alpha=1}^{N}(\underline{x}^{\alpha}-\beta_1^{o} z^{\alpha}_{(1)} -\hat{\hat{\beta}}_2 z^{\alpha}_{(2)})(\underline{x}^{\alpha}-\beta_1^{o} z^{\alpha}_{(1)}-\hat{\hat{\beta}}_2 z^{\alpha}_{(2)})' \tag{7.47}$$

Lemma 7.4

$$N\hat{\hat{\Sigma}} = N\hat{\Sigma} + (\hat{\beta}_1-\beta_1^{o})(A_{(11)}-A_{(12)}A^{-1}_{(22)}A_{(21)})(\hat{\beta}_1-\beta_1^{o})'.$$

Proof.

$$C = \hat{\beta}A = (\hat{\beta}_1, \hat{\beta}_2)\begin{pmatrix} A_{(11)}, A_{(12)} \\ A_{(21)}, A_{(22)} \end{pmatrix} = (\hat{\beta}_{(1)}A_{(11)} + \hat{\beta}_{(2)}A_{(21)}, \hat{\beta}_1 A_{(12)} + \hat{\beta}_2 A_{(22)})$$

we obtain

$$C_2 = \hat{\beta}_{(1)}A_{(12)} + \hat{\beta}_{(2)}A_{(22)}.$$

Since A is positive definite (from the rank consideration), $A^{-1}_{(22)}$ exists and

$$\hat{\beta}_2 = C_2 A^{-1}_{(22)} - \hat{\beta}_{(1)}A_{(12)}A^{-1}_{(22)}, \quad \hat{\beta}_2 - \hat{\hat{\beta}}_2 = -(\hat{\beta}_1 - \beta_1^{o})A_{(12)}A^{-1}_{(22)}. \tag{7.48}$$

Now, under H_o, using (7.47) - (7.48)

$$\begin{aligned}(x-\beta z) &= (x-\hat{\beta}z) + (\hat{\beta}_2-\beta_2)z_2 + (\hat{\beta}_1-\beta_1^{o})z_1 \\ &= (x-\hat{\beta}z) + (\hat{\hat{\beta}}_2-\beta_2)z_2 + (\hat{\beta}_1-\beta_1^{o})(z_1-A_{(12)}A^{-1}_{(22)}\ z_2).\end{aligned} \tag{7.49}$$

Now

$$(z_1-A_{(12)}A^{-1}_{(22)}z_2)z_2' = A_{(12)}-A_{(12)} = 0,$$

$$(x-\hat{\beta}z)z' = xz'-xz'(zz')^{-1}zz' = xz' - xz' = 0. \tag{7.50}$$

Since (7.49) implies that

$$(x-\hat{\beta}z)z_i' = 0, \quad i = 1, 2,$$

we obtain from (7.49) - (7.50)

$$\begin{aligned}(x-\beta z)(x-\beta z)' &= (x-\hat{\beta}z)(x-\hat{\beta}z)' + (\hat{\hat{\beta}}_2-\beta_2)A_{(22)}(\hat{\hat{\beta}}_2-\beta_2)' \\ &\quad + (\hat{\beta}_1-\beta_1^{o})(A_{(11)}-A_{(12)}A^{-1}_{(22)}\ A_{(21)})(\hat{\beta}_1-\beta_1^{o})',\end{aligned}$$

$$(x-\beta_1^{o}z_1-\hat{\hat{\beta}}_2z_2) = (x-\hat{\beta}z) + (\hat{\beta}_1-\beta_1^{o})(z_1-A_{(12)}A^{-1}_{(22)}\ z_2).$$

Thus

$$N\hat{\hat{\Sigma}} = (x-\beta_1^{o}z_1-\hat{\hat{\beta}}_2z_2)(x-\beta_1^{o}z_1-\hat{\hat{\beta}}_2z_2)'$$

$$=(x-\hat{\beta}z)(x-\hat{\beta}z)' + (\hat{\beta}_1-\beta_1^{0})(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})(\hat{\beta}_1-\beta_1^{0})'$$

Q.E.D.

Using (7.40), (7.47) and Lemma 7.4 we conclude that the likelihood ratio test of H_0: $\beta_1 = \beta_1^{0}$ rejects whenever

$$u = \frac{\det\left(\sum_{\alpha=1}^{N} (\underline{x}^{\alpha}-\hat{\beta}\underline{z}^{\alpha})(\underline{x}^{\alpha}-\hat{\beta}\underline{z}^{\alpha})'\right)}{\det\left(\sum_{\alpha=1}^{N} (\underline{x}^{\alpha}-\hat{\beta}\underline{z}^{\alpha})(\underline{x}^{\alpha}-\hat{\beta}\underline{z}^{\alpha})'+(\hat{\beta}_1-\beta_1^{0})(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})(\hat{\beta}_1-\beta_1^{0})'\right)}$$

$$\geq c \tag{7.51}$$

where the constant c depends on the level of significance α of the test. We shall now show that the corresponding statistic U in (7.51) is distributed as the statistic U in (7.9). Wilks (1932) first derived the likelihood test criterion for the special case of testing the equality of mean vectors of several populations. Wilks (1934) and Bartlett (1934) extended its use to multivariate regression model. In what follows for notational simplicity we shall use the same notation for the estimate and the ~~correspond~~ corresponding estimator.

For the maximum likelihood estimator $\hat{\beta}$ of β

$$E(\hat{\beta}) = E\left(\sum_{\alpha=1}^{N} \underline{x}^{\alpha}\underline{z}^{\alpha'}A^{-1}\right)=\beta AA^{-1} = \beta . \tag{7.52}$$

The covariance between the ith row vector $\hat{\underline{\beta}}_i$ and the jth row vector $\hat{\underline{\beta}}_j$ of $\hat{\beta}$ is

$$E(\hat{\underline{\beta}}_i - \underline{\beta}_i)(\underline{\beta}_j - \underline{\beta}_j)'$$

$$= A^{-1}E\left\{\sum_{\alpha=1}^{N} (X_{\alpha i}-E(X_{\alpha i}))\underline{z}^{\alpha}\left(\sum_{\gamma=1}^{N} (X_{\gamma j}-E(X_{\gamma j}))\underline{z}^{\gamma'}\right\}A^{-1}\right.$$

$$= A^{-1} \sum_{\alpha=1}^{N} \sigma_{ij} \underline{z}^{\alpha} \underline{z}^{\alpha'} A^{-1} = \sigma_{ij} A^{-1} ,$$

(7.53)

where $\Sigma = (\sigma_{ij})$. Obviously thus the row vectors $\hat{\beta}_1, \ldots, \hat{\beta}_p$ are normally distributed with means $\beta_1, \ldots, \beta_p$ and covariance matrix

$$\begin{pmatrix} \sigma_{11} A^{-1}, \ldots, & \sigma_{1p} A^{-1} \\ \sigma_{21} A^{-1}, \ldots, & \sigma_{2p} A^{-1} \\ \cdot\cdot \quad , \ldots, & \cdot\cdot \\ \sigma_{p1} A^{-1}, \ldots, & \sigma_{pp} A^{-1} \end{pmatrix} = \Sigma \times A^{-1} \tag{7.54}$$

Theorem 7.5. $N \hat{\Sigma} = \sum_{\alpha=1}^{N} \underline{X}^{\alpha} \underline{X}^{\alpha'} - \hat{\beta} A \hat{\beta}'$ is distributed independently of $\hat{\beta}$ as $W_p(\Sigma, N-s)$.

Proof. Let F be an $s \times s$ nonsingular matrix such that $FAF' = I$ and let $E_2 = Fz$. Then $E_2 E_2' = Fzz'F' = FAF' = I$. This implies that the s rows of E_2 are orthogonal vectors of unit length. Thus it is possible to find an $(N-s) \times N$ matrix E_1 such that $E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}$ is an orthogonal matrix. Let $Y = (\underline{Y}^1, \ldots, \underline{Y}^N) = XE'$. Thus the column vectors of Y are independently distributed normal vectors with the same covariance matrix Σ and

$$E(Y) = \beta z E' = \beta F^{-1} E_2 (E_1', E_2') = (0, \beta F^{-1}).$$

Since

$$XX' = \sum_{\alpha=1}^{N} \underline{X}^{\alpha} \underline{X}^{\alpha'} = YY' = \sum_{\alpha=1}^{N} \underline{Y}^{\alpha} \underline{Y}^{\alpha'} , \tag{7.55}$$

$$\hat{\beta} A \hat{\beta}' = Xz'A^{-1}AA^{-1}zX' = YEE_2' (F^{-1})' A^{-1} F^{-1} E_2 E' Y'$$

$$= Y \begin{pmatrix} 0 \\ I \end{pmatrix} (0, I) Y' = \sum_{\alpha=N-s+1}^{N} \underline{Y}^{\alpha} \underline{Y}^{\alpha'} \tag{7.56}$$

we obtain

$$N\hat{\Sigma} = \sum_{\alpha=1}^{N-s} \underline{Y}^{\alpha}\underline{Y}^{\alpha'} . \qquad (7.57)$$

where $\underline{Y}^{\alpha}, \alpha = 1,\dots,N-s$ are independently distributed normal p-vectors with means $\underline{0}$ and the same covariance matrix Σ. Thus from (7.56)-(7.57) $N\hat{\Sigma}$ is distributed as $W_p(N-s,\Sigma)$ independently of $\hat{\beta}$. Q.E.D.

Theorem 7.6. Under H_0 $(\hat{\beta}_1-\beta_1^{o})(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})(\hat{\beta}_1-\beta_1^{o})'$ is distributed as $W_p(r,\Sigma)$ (independently of $N\hat{\Sigma}$).

Proof. From (7.54) the covariance between the ith and jth rows of the estimator $\hat{\beta}_1$ is $\delta_{ij}\,(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})^{-1}$. Let B be an r × r nonsingular matrix such that

$$B(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})B' = I$$

and let $\hat{\beta}_1-\beta_1^{o} = YB = (Y^1,\dots,Y^r)\,B$. Then

$$(\hat{\underline{\beta}}_1-\beta_1^{o})(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})(\hat{\beta}_1-\beta_1^{o})' = \sum_{\alpha=1}^{r}\underline{Y}^{\alpha}\underline{Y}^{\alpha'} = YY' .$$

Under H_0, since $E(\hat{\beta}_1) = \beta_1^{o}$,

$$E(Y) = E((\hat{\beta}_1-\beta_1^{o})B^{-1}) = 0.$$

Let the ith and jth row of Y be $\underline{Y}_i$, $\underline{Y}_j$ respectively and let the ith and the jth row of $\hat{\beta}_1$ be $\hat{\underline{\beta}}_{i1}$, $\hat{\underline{\beta}}_{j1}$ respectively. Then

$$E(\underline{Y}_i'\,\underline{Y}_j) = E((B^{-1})'\,(\hat{\underline{\beta}}_{i1}-\underline{\beta}_{i1}^{o})'(\hat{\underline{\beta}}_{j1}-\underline{\beta}_{j1})B^{-1})$$

$$=\sigma_{ij}(B^{-1})'\,(A_{(11)}-A_{(12)}\,A_{(22)}^{-1}\,A_{(21)})^{-1}B^{-1}$$

$$=\sigma_{ij}(B(A_{(11)}-A_{(12)}A_{(22)}^{-1}A_{(21)})B')^{-1}$$

$$=\sigma_{ij}\; I.$$

Thus $\sum_{\alpha=1}^{r}\underline{Y}^{\alpha}\underline{Y}^{\alpha'}$ is distributed as $W_p(\Sigma,r)$ $(r\geq p)$ when H_0 is true. The independence of $\hat{\Sigma}$ follows from Theorem 7.6. Q.E.D.

7.6. APPLICATIONS OF MULTIVARIATE GENERAL LINEAR HYPOTHESIS

Here a fixed set of ν treatments, experimented conditions or diagonistic classifications, etc. are of interest to the experimenter and in the presence of each some p response variates will be recorded on the sampling units. These measurements will be assumed to be independent observations on ν p-dimensional normal populations with different means $\underline{\mu}^{\alpha}$, $\alpha = 1,\ldots,\nu$ but with the same positive definite covariance matrix Σ . Let $\underline{x}_{ij} = (x_{ij1},\ldots,x_{ijp})'$, $j = 1,\ldots,N_i$, $i = 1,\ldots,\nu$ $(N_i > p)$ be a sample of size N_i from the ith normal population with mean $\underline{\mu}^i$ and covariance matrix Σ (positive definite). We are interested here in testing the hypothesis

$$H_o\colon\ \underline{\mu}^1 = \ldots = \underline{\mu}^{\nu} = \underline{\mu}\ \text{(say)}.$$

In the notation of Section 7.1 it is a multivariate linear hypothesis with $s-r = 1$, $N = \sum_{i=1}^{\nu} N_i$, $s = \nu$. Define

$$N_i\underline{x}_{i.} = \sum_j \underline{x}_{ij},\quad N\underline{x}_{..} = \sum_i\sum_j \underline{x}_{ij},\quad s_i = \sum_j (\underline{x}_{ij} - \underline{x}_{i.})(\underline{x}_{ij} - \underline{x}_{i.})' . \tag{7.58}$$

$$s = \sum_i s_i .$$

The likelihood of the observation vectors $\underline{x}_{ij}$, $j = 1,\ldots,N_i$, $i = 1,\ldots,\nu$ is

$$L(\underline{\mu}^1,\ldots,\underline{\mu}^{\nu}, \Sigma) = (2\pi)^{-Np/2}(\det\Sigma\ )^{-N/2} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(\sum_i\sum_j(\underline{x}_{ij}-\underline{\mu}^i)(\underline{x}_{ij}-\underline{\mu}^i)')\}. \tag{7.59}$$

The parametric space Ω is given by

$$\Omega = \{(\underline{\mu}^1,\ldots,\underline{\mu}^{\nu},\Sigma)\}$$

and it reduces to $\omega = \{(\underline{\mu},\Sigma)\}$ when H_o is true. Using Lemma 7.1

$$\max_{\Omega} L(\underline{\mu}^1,\ldots,\underline{\mu}^{\nu},\Sigma) = \max_{\Sigma}\ (2\pi)^{-Np/2}(\det\Sigma)^{-N/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}s\}$$
$$= (\frac{2\pi}{N})^{-Np/2}(\det s)^{-N/2} \exp\{-Np/2\}. \tag{7.60}$$

Under H_o the likelihood function reduces to $L(\underline{\mu},\Sigma)$ and

$$\max_{\omega} L(\underline{\mu},\Sigma) = \max_{\Sigma} (2\pi)^{-Np/2}(\det\Sigma)^{-N/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{..}))\}$$

$$=\left(\frac{2\pi}{N}\right)^{-Np/2}\left(\det\left(\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{..})'\right)\right)^{-N/2}\exp\{-Np/2\}. \qquad (7.61)$$

Since

$$\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{..})'$$

$$=\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{i.})(\underline{x}_{ij}-\underline{x}_{..})' + \sum_i N_i(\underline{x}_{i.}-\underline{x}_{..})(\underline{x}_{i.}-\underline{x}_{..})'$$

we obtain from (7.60)-(7.61)

$$\lambda = \frac{\max_{\omega} L(\underline{\mu},\Sigma)}{\max_{\Omega} L(\underline{\mu}',\ldots,\underline{\mu}^{\nu},\Sigma)} = \left[\frac{\det s}{\det(s+\sum_i N_i(\underline{x}_{i.}-\underline{x}_{..})(\underline{x}_{i.}-\underline{x}_{..})')}\right]^{N/2}$$

$$=\left[\frac{\det(s)}{\det(s+a)}\right]^{N/2}$$

where $a = \sum_i N_i(\underline{x}_{i.}-\underline{x}_{..})(\underline{x}_{i.}-\underline{x}_{..})'$. Thus the likelihood ratio test of H_o rejects H_o whenever

$$u = \frac{\det s}{\det(s+a)} \le k \qquad (7.62)$$

where k is a constant depending on the level of significance α of the test. Though from Section 7.1 it follows that under H_o the statistic U (with values u) is distributed as $U_{p,\nu-1,N-\nu}$, we shall present below a direct proof of this as it will help us to develop a better understanding of various distributions involved in two and higher-way multivariate classifications.

From Theorem 2.12, A (being a function of $\underline{X}_{i.}, i = 1,\ldots,\nu$) and S (with values s) are independent. Let

$$X = (\underline{X}_{11},\dots,\underline{X}_{1N_1},\dots,\underline{X}_{\nu 1},\dots,\underline{X}_{\nu N_\nu})$$

and let the matrices A_1, A_2 and A_3, each of the dimension $N \times N$ be given by

$$A_3 = \begin{pmatrix} \frac{1}{N},\dots,\frac{1}{N} \\ \cdot \;\cdots\; \cdot \\ \frac{1}{N},\dots,\frac{1}{N} \end{pmatrix}, \quad A_2 = B - A_3, \quad A_1 = I-B$$

where B is a diagonal matrix in the block form with diagonal elements $B_{(11)},\dots,B_{(\nu\nu)}$, $B_{(ii)}$ being a $N_i \times N_i$ matrix each of whose elements equal to $1/N_i$ and I is the $N \times N$ identity matrix. Obviously $A_iA_i = A_i$, $i = 1,2,3$. In other words A_1, A_2 and A_3 are idempotent matrices. Thus

$$\text{rank}(A_3) = \text{tr}\, A_3 = 1, \quad \text{rank}(A_2) = \text{tr}\, B - \text{tr}\, A_3 = \nu-1,$$

$$\text{rank}(A_1) = \text{tr}\, I - \text{tr}\, B = N-\nu .$$

Furthermore $A_iA_j = 0$ $(i \neq j)$. Now

$$\sum_i \sum_j (\underline{X}_{ij} - \underline{X}_{i.})(\underline{X}_{ij}-\underline{X}_{i.})' = XA_1X', \quad \sum_i N_i (\underline{X}_{i.}-\underline{X}_{..})(\underline{X}_{i.}-\underline{X}_{..})'$$

$$= XA_2X'.$$

Since A_1 is idempotent of rank $N-\nu$ there exists an orthogonal matrix F_1 such that $F_1'A_1F_1 = \begin{pmatrix} I(N-\nu), 0 \\ 0 \quad 0 \end{pmatrix}$ where $I(\nu-1)$ is the $(\nu-1) \times (\nu-1)$ identity matrix. Hence

$$XA_1X' = (XF_1)(F_1'A_1F_1)(XF_1)' = \sum_{\alpha=1}^{N-\nu} \underline{Y}^{\alpha}\underline{Y}^{\alpha'}$$

where $(XF_1) = (\underline{Y}^1,\dots,\underline{Y}^N) = Y$ and $\underline{Y}^{\alpha} = (Y_{\alpha 1},\dots,Y_{\alpha p})'$, $\alpha = 1,\dots,N$.

As

$$XA_1 = (\underline{X}_{11}-\underline{X}_{1.},\dots,X_{1N_1}-\underline{X}_{1.},\dots,\underline{X}_{\nu 1}-\underline{X}_{\nu .},\dots,\underline{X}_{\nu N_\nu}-\underline{X}_{\nu .})$$

we obtain $E(XA_1) = 0$. But

$$E(XA_1) = E(YF_1'A_1) = E(Y)F_1'A_1 = 0$$

which implies that $E(Y)\ F_1'A_1F_1 = ((E\underline{Y}^1),\ldots E(\underline{Y}^{N-\nu}),\ \underline{0},\ldots,\underline{0})=0$. Since F_1 is an orthogonal matrix $\underline{Y}^1,\ldots,\underline{Y}^{N-\nu}$ are independently distributed normal p-vectors with mean $\underline{0}$ and the same covariance matrix Σ. Hence XAX' is distributed as $W_p(\Sigma, N-\nu)$. Similarly since A_2 is an idempotent matrix of rank $\nu-1$,

$$XA_2=(\underline{X}_{1.}-\underline{X}_{..},\ldots,\underline{X}_{1.}-\underline{X}_{..},\ldots,\underline{X}_{\nu.}-\underline{X}_{..},\ldots,\underline{X}_{\nu.}-\underline{X}_{..})$$

and $E(XA_2) = 0$ under H_o, we conclude that XA_2X' is distributed as $W_p(\Sigma,\nu-1)$ when H_o is true. Thus under H_o $U = \det S/\det(S+A)$ is distributed $U_{p,\nu-1,N-\nu}$ and we have discussed its distribution in Section 7.3.

In the setting of multivariate regression model the mathematical model of the multivariate one-way classification can be written as

$$E(\underline{X}_{ij}) = \beta\underline{z}^{ij},\ j = 1,\ldots,N_i,\ i = 1,\ldots,\nu,$$

$$\mathrm{Cov}(\underline{X}_{ij}) = \Sigma \text{ (positive definite)}$$

or equivalently:

$$E(X) = \beta z,\ \mathrm{Cov}(\underline{X}_{ij}) = \Sigma \text{ (positive definite)}, \tag{7.63}$$

where β is a $p \times \nu$ matrix of unknown elements, z is a $\nu \times N$ matrix of known elements and are given by

$$\mu = (\underline{\mu}^1-\underline{\mu}^\nu,\ \underline{\mu}^2-\underline{\mu}^\nu,\ldots,\underline{\mu}^{\nu-1}-\underline{\mu}^\nu,\underline{\mu}^\nu),$$

$$z = (\underline{z}_{11},\ldots,\underline{z}_{1N_1},\ldots,\underline{z}_{\nu1},\ldots,\underline{z}_{\nu N_\nu}),$$

$$= \begin{pmatrix} 1 & \cdot\cdot & 1 & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 \\ 0 & \cdot\cdot & 0 & 1 & \cdot\cdot & 1 & \cdot\cdot & 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 \\ 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 & \cdot\cdot & 1 & \cdot\cdot & 1 & 0 & \cdot\cdot & 0 \\ 1 & \cdot\cdot & 1 & 1 & \cdot\cdot & 1 & \cdot\cdot & 1 & \cdot\cdot & 1 & 1 & \cdot\cdot & 1 \end{pmatrix}$$

with $\underline{z}_{ij} = (z_{ij1},\ldots,z_{ij\nu})'$, $j = 1,\ldots,N_i$, $i = 1,\ldots,\nu$.

The hypothesis H_o reduces to $H_o:\beta_1 = 0$ where $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\underline{\mu}^1 - \underline{\mu}^\nu, \ldots, \underline{\mu}^{\nu-1} - \underline{\mu}^\nu)$.

Note that this representation is not the only possible in this case. One other natural representation is $\beta = (\underline{\mu}^1 - \underline{\mu}_{.}, \ldots, \underline{\mu}^\nu - \underline{\mu}_{.})$ with $\nu\underline{\mu} = \sum_i \underline{\mu}^i \underline{\mu}$ and the hypothesis $H_o^{\cdot}$ in this case reduces to $\beta = 0$.

The likelihood of the observation x on X is given by

$$L(\beta,\Sigma) = (2\pi)^{-Np/2}(\det \Sigma)^{-N/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(x-\beta z)(x-\beta z)'\} \tag{7.64}$$

Since the rank of the $\nu \times N$ matrix $z\ (\nu < N)$ is ν, $(zz')^{-1}$ exists and the unique maximum likelihood estimate of β is

$$\hat{\beta} = xz'(zz')^{-1} = (\underline{x}_{1.}-\underline{x}_{\nu.}, \underline{x}_{2.}-\underline{x}_{\nu.}, \ldots, \underline{x}_{\nu-1.}-\underline{x}_{\nu.}, \underline{x}_{\nu.}). \tag{7.65}$$

From (7.64) using Lemma (7.1) the maximum likelihood estimate of Σ is

$$\hat{\Sigma} = \frac{1}{N}(xx' - \hat{\beta}(zz)\hat{\beta}') = \frac{1}{N}\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{i.})(\underline{x}_{ij}-\underline{x}_{i.})' \tag{7.66}$$

Under H_o the maximum likelihood estimate of Σ is given by

$$\hat{\hat{\Sigma}} = \sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{..})' = \hat{\Sigma} + \sum_i N_i(\underline{x}_{i.}-\underline{x}_{..})(\underline{x}_{i.}-\underline{x}_{..})' \tag{7.67}$$

From (7.66)-(7.67) the likelihood ratio test of H_o rejects H_o whenever

$$u = \frac{\det(s)}{\det(a+s)} \leq k \tag{7.68}$$

where k is a constant depending on the level of significance of the test and the corresponding statistic U is distributed as $U_{p,\nu-1,N-\nu}$ when H_o is true.

As discussed in Section 7.1.1 other invariant tests of H_o are as follows.

(i) Lawley-Hotelling's test criterion: Reject H_o whenever $\operatorname{tr}(as^{-1}) \geq k$.

(ii) Pillai's test criterion: Reject H_0 whenever $\mathrm{tr}(a(a+s)^{-1} \geq k$.

(iii) Roy's test criteria based on the maximum and the minimum characteristic roots of $a(a+s)^{-1}$.

If we reject H_0 by using any one of the four test criteria discussed above we conclude that not all the mean vectors are equal. Naturally such rejection leads to further investigation to decide which mean vectors are different.

Since $\sqrt{\frac{N_iN_j}{N_i+N_j}}(\underline{X}_{i.}-\underline{X}_{j.})$ has a p-variate normal distribution with mean $\sqrt{\frac{N_iN_j}{N_i+N_j}}(\underline{\mu}^i-\underline{\mu}^j)$ and positive definite covariance matrix Σ and the random matrix S is distributed independently of $\sqrt{\frac{N_iN_j}{N_i+N_j}}(\underline{X}_{i.}-\underline{X}_{j.})$ as $W_p(\Sigma,N-\nu)$, $T^2 = \frac{N_iN_j}{N_i+N_j}(\underline{X}_{i.}-\underline{X}_{j.})' S^{-1}(\underline{X}_{i.}-\underline{X}_{j.})$ is distributed as Hotelling's T^2 statistic with $N-\nu$ degrees of freedom and noncentrality parameter $\frac{N_iN_j}{N_i+N_j}(\underline{\mu}^i-\underline{\mu}^j)' \Sigma^{-1}(\underline{\mu}^i-\underline{\mu}^j)$. Thus from chapter 2 the optimum test of level α of the null hypothesis $\underline{\mu}^i = \underline{\mu}^j$ rejects it whenever

$$(N-\nu)\left(\frac{N_iN_j}{N_i+N_j}\right)(\underline{x}_{i.}-\underline{x}_{j.})' s^{-1}(\underline{x}_{i.}-\underline{x}_{j.}) \geq \frac{(N-\nu)p}{(N-\nu-p+1)} F_{1-\alpha,p,N-\nu-p+1}.$$

Note that from Theorem 2.14 under the null hypothesis $\frac{(N-\nu-p+1)}{p(N-\nu)}T^2$ has central F-distribution with $(p,N-\nu-p+1)$ degrees of freedom. Thus on the basis of such comparisons for all possible pairs $(\underline{\mu}^i,\underline{\mu}^j)(i\neq j)$ it will be possible to divide the ν populations into different groups such that the populations of the same group have the same mean and those of different groups have different means.

7.6.1. Experimental Design Model of the Multivariate One-way Classifications: We have discussed in complete details the univariate one-way classification in Section 3.6. We shall consider here a formulation of the multivariate one-way classification in terms of multivariate regression model where the design matrix z does not have full rank. Consider the multivariate regression model (instead of (7.63)).

$$E(X) = \beta z, \quad \mathrm{Cov}(\underline{X}_{ij}) = \Sigma \text{ (positive definite)}$$

$$\text{where } \beta = (\mu, \underline{\alpha}^1, \ldots, \underline{\alpha}^\nu) \text{ and} \tag{7.69}$$

$$z = \begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix}$$

The matrix Σ is of dimension $p \times (\nu+1)$ and the matrix z is of dimension $(\nu+1) \times N$. Obviously the rank of z is ν (assuming $\nu+1 < N$).

Clearly (7.69) can be obtained from (7.63) writing $\underline{\mu} = \nu^{-1} \sum_i \underline{\mu}^i$, $\underline{\alpha}^i = \underline{\mu}^i - \underline{\mu}$, $i = 1, \ldots, \nu$ and in this obviously $\sum_i \underline{\alpha}^i = \underline{0}$.

As in the univariate case all linear combinations of the $\underline{\alpha}^i$ are not estimable. To see what kind of linear combinations of the $\underline{\alpha}^i$ are estimable let us consider the estimator $\sum_i \sum_j a_{ij} \underline{X}_{ij}$ where the a_{ij} are reals with $\sum_j a_{ij} = c_i$ (say). Since

$$E\Big(\sum_i \sum_j a_{ij} \underline{X}_{ij}\Big) = \sum_i \sum_j a_{ij} (\underline{\mu} + \underline{\alpha}^i) = \sum_i c_i \underline{\alpha}^i + \underline{\mu}\Big(\sum_i c_i\Big)$$

we conclude that any linear combination $\sum_i c_i \underline{\alpha}^i$ is estimable if $\sum_i c_i = 0$. In other words $\sum_i c_i \underline{\alpha}^i$ is a contrast. In particular all elementary contrasts $\underline{\alpha}^i - \underline{\alpha}^j$ $(i \neq j)$ are estimable. But neither $\sum_i \underline{\alpha}^i$ nor any one of the parameters $\underline{\mu}, \underline{\alpha}^1, \ldots, \underline{\alpha}^\nu$ is estimable. However, if we impose the side restriction

$\sum_i \hat{\underline{\alpha}}^i = \underline{0}$ then the unique estimate of $\underline{\mu}, \underline{\alpha}^i$, $i = 1,\dots,\nu$ are

$$\hat{\underline{\mu}} = \underline{x}_{..}, \quad \hat{\underline{\alpha}}^i = \underline{x}_{i.} - \underline{x}_{..}, \quad i = 1,\dots,\nu. \tag{7.70}$$

Thus for the model (7.69) with the added restriction that $\sum_i \underline{\alpha}^i = \underline{0}$ the unique estimates of $\underline{\mu}$ and the $\underline{\alpha}^i$ are given by (7.70). In this case the hypothesis $H_o: \underline{\alpha}^1 = \dots = \underline{\alpha}^\nu$ reduces to H_o: all $\underline{\alpha}^i = \underline{0}$ and the likelihood ratio test (Wilks' criterion) of H_o is given by (7.68).

If H_o is rejected it is necessary to test estimable hypothesis of the form $H_o^*: \sum_i c_i \underline{\alpha}^i = \underline{\delta}$ where $\sum_i c_i = 0$ and $\underline{\delta}$ is a fixed vector and the appropriate level α invariant test of H_o^* is given by,

Reject H_o^* whenever,

$$(N-\nu)\ \lambda\ \Big(\sum_i c_i \underline{x}_{i.} - \underline{\delta}\Big)' s^{-1} \Big(\sum_i c_i \underline{x}_{i.} - \underline{\delta}\Big) \geq \frac{p(N-\nu)}{(N-\nu-p+1)} F_{1-\nu,p,N-\nu-p+1},$$

where

$\lambda^{-1} = \sum_i (c_i{}^2/N_i)$. Note that under H_o^*

$$(N-\nu)\ \lambda\Big(\sum_i c_i \underline{X}_{i.} - \underline{\delta}\Big)' S^{-1}\Big(\sum_i c_i \underline{X}_{i.} - \underline{\delta}\Big)$$

is distributed as Hotelling's T^2-statistic with $N-\nu$ degrees of freedom.

7.6.2. Multivariate Two-way Classifications with Single Observations Per Cell:

The experimental set-up is the same as the univariate two-way classification as discussed in Chapter 3. Suppose that we have a set of independently distributed normal p-vectors $\underline{X}_{ij} = (X_{ij1},\dots,X_{ijp})'$, $i = 1,\dots,I$, $j = 1,\dots,J$ with

$$E(\underline{X}_{ij}) = \underline{\mu} + \underline{\alpha}^i + \underline{\beta}^j, \quad \mathrm{Cov}(\underline{X}_{ij}) = \Sigma \tag{7.71}$$

where Σ is a positive definite matrix, $\underline{\mu} = (\mu_1,\dots,\mu_p)'$, $\underline{\alpha}^i = (\alpha_{i1},\dots,\alpha_{ip})'$ satisfying $\sum_i \underline{\alpha}^i = \underline{0}$ and $\underline{\beta}^j = (\beta_{j1},\dots,\beta_{jp})'$ satisfying $\sum_j \underline{\beta}^j = \underline{0}$. With this model the hypotheses of interest are

$$H_{\underline{\alpha}} : \text{all } \underline{\alpha}^i = \underline{0}, \quad H_{\underline{\beta}} : \text{all } \underline{\beta}^j = \underline{0}.$$

In the terminology of Section 7.1 $H_{\underline{\alpha}}$ is a multivariate linear hypothesis with $r = I-1$, $s = I+J-1$ and $N-s = (I-1)(J-1)$. Similarly $H_{\underline{\beta}}$ is a multivariate linear hypothesis with $r = J-1$, $s = I+J-1$ and $N-s = (I-1)(J-1)$. The likelihood of the observations $\underline{x}_{ij}$ is given by

$$L(\underline{\mu},\underline{\alpha}^1,\ldots,\underline{\alpha}^I,\underline{\beta}^1,\ldots,\underline{\beta}^J,\Sigma)$$
$$=(2\pi)^{-IJp/2}(\det\Sigma)^{-IJ/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(\sum_i\sum_j(\underline{x}_{ij}-\underline{\mu}-\underline{\alpha}^i-\underline{\beta}^j)$$
$$\times(\underline{x}_{ij}-\underline{\mu}-\underline{\alpha}^i-\underline{\beta}^j)')\}. \tag{7.72}$$

Under π_Ω the maximum likelihood estimates of $\underline{\mu},\underline{\alpha}^i,\underline{\beta}^j$ are $\hat{\underline{\mu}} = \underline{x}_{..}$, $\hat{\underline{\alpha}}^i = \underline{x}_{i.}-\underline{x}_{..}$, $\hat{\underline{\beta}}^j = \underline{x}_{.j}-\underline{x}_{..}$. By using Lemma 7.1

$$\max_{\Omega} L(\underline{\mu},\underline{\alpha}^1,\ldots,\underline{\alpha}^I,\underline{\beta}^1,\ldots,\underline{\beta}^J,\Sigma)$$
$$=\max_{\Sigma}(2\pi)^{-IJp/2}(\det\Sigma)^{-IJ/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{i.}-\underline{x}_{.j}+\underline{x}_{..})$$
$$\times(\underline{x}_{ij}-\underline{x}_{i.}\ \underline{x}_{.j}+\underline{x}_{..})')\}.$$
$$=\left(\frac{2\pi}{IJ}\right)^{-IJp/2}[\det(\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{i.}-\underline{x}_{.j}+\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{i.}-\underline{x}_{.j}+\underline{x}_{..})')]^{-IJ/2}$$
$$\exp\{-IJp/2\}. \tag{7.73}$$

Under $H_{\underline{\alpha}}$ the maximum likelihood estimate of $\underline{\mu}$, $\underline{\beta}^j$ are $\hat{\hat{\underline{\mu}}}=\underline{x}_{..}$, $\hat{\hat{\underline{\beta}}}^j = \underline{x}_{.j}-\underline{x}_{..}$ and

$$\max_{H_{\underline{\alpha}}} L(\underline{\mu},\underline{\beta}^1,\ldots,\underline{\beta}^J,\Sigma)$$
$$=\left(\frac{2\pi}{IJ}\right)^{-IJp/2}[\det(\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{.j})(\underline{x}_{ij}-\underline{x}_{.j})')]^{-IJ/2} \tag{7.74}$$
$$\exp\{-IJp/2\}.$$

Since

$$\sum_i\sum_j(\underline{x}_{ij}-\underline{x}_{.j})(\underline{x}_{ij}-\underline{x}_{.j})' = c + a$$

where

$$c = \sum_i \sum_j (\underline{x}_{ij}-\underline{x}_{i.}-\underline{x}_{.j}+\underline{x}_{..})(\underline{x}_{ij}-\underline{x}_{i.}-\underline{x}_{.j}+\underline{x}_{..})'$$

$$a = \sum_i J(\underline{x}_{i.}-\underline{x}_{..})(\underline{x}_{i.}-\underline{x}_{..})',$$

the likelihood ratio test of $H_{\underline{\alpha}}$ rejects $H_{\underline{\alpha}}$ whenever

$$u_1 = \frac{\det c}{\det(a+c)} \le k$$

where the constant k depends on the level of significance α of the test. Under $H_{\underline{\alpha}}$ the statistic U_1 (with values u_1) is distributed as $U_{p,I-1,(I-1)(J-1)}$. As in the case of one-way classification it can also be verified directly that the random matrix $C = \sum_i \sum_j (\underline{X}_{ij}-\underline{X}_{i.}-\underline{X}_{.j}+\underline{X}_{..})(\underline{X}_{ij}-\underline{X}_{i.}-\underline{X}_{.j}+\underline{X}_{..})'$ is distributed independently of $A = \sum_i J(\underline{X}_{i.}-\underline{X}_{..})(\underline{X}_{i.}-\underline{X}_{..})'$ as $W_p(\Sigma,(I-1)(J-1))$ and under $H_{\underline{\alpha}}$ A is distributed $W_p(\Sigma, I-1)$. Similarly it may be verified that the likelihood ratio test of $H_{\underline{\beta}}$ rejects $H_{\underline{\beta}}$ whenever

$$u_2 = \frac{\det c}{\det(b+c)} \le k$$

where the constant k depends on the level of significance α of the test and $b = \sum_j I(\underline{x}_{.j}-\underline{x}_{..})(\underline{x}_{.j}-\underline{x}_{..})'$. The statistic U_2 (with values u_2) is distributed as $U_{p,J-1,(I-1)(J-1)}$. The Wishart matrix C is independent of $B = \sum_j I(\underline{X}_{.j}-\underline{X}_{..})(\underline{X}_{.j}-\underline{X}_{..})'$.

In order that C,B,A are positive definite with probability one it is necessary to have $p \le I-1$ and $p \le J-1$. Other invariant tests of $H_{\underline{\alpha}}$ and $H_{\underline{\beta}}$ can be obtained following (7.14)-(7.17). In terms of the multivariate regression model the mathematical model of the multivariate two-way classification can be written as

$$E(X) = \beta z, \quad \mathrm{Cov}(\underline{X}_{ij}) = \Sigma \quad \text{(positive definite)} \tag{7.75}$$

where $X = (\underline{X}_{11},\ldots,\underline{X}_{1J},\ldots,\underline{X}_{I1},\ldots,\underline{X}_{IJ})$ is a $p \times IJ$ matrix, $\beta = (\underline{\mu},\underline{\alpha}^1,\ldots,\underline{\alpha}^I, \underline{\beta}^1,\ldots,\underline{\beta}^J)$ is a $p \times (I+J+1)$ matrix and $z = (\underline{z}_{11},\ldots,\underline{z}_{1J},\ldots,\underline{z}_{I1},\ldots,\underline{z}_{IJ})$

$$= \begin{pmatrix} 1 & \cdot\cdot & 1 & 1 & \cdot\cdot & 1 & \cdot\cdot & 1 & \cdot\cdot & 1 \\ 1 & \cdot\cdot & 1 & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 \\ 0 & \cdot\cdot & 0 & 1 & \cdot\cdot & 1 & \cdot\cdot & 0 & \cdot\cdot & 0 \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot \\ 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 & \cdot\cdot & 1 & \cdot\cdot & 1 \\ 1 & \cdot\cdot & 0 & 1 & \cdot\cdot & 0 & \cdot\cdot & 1 & \cdot\cdot & 0 \\ 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 & \cdot\cdot & 0 \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot \\ 0 & \cdot\cdot & 0 & 0 & \cdot\cdot & 0 & \cdot\cdot & \cdot & \cdot\cdot & \cdot \\ 0 & \cdot\cdot & 1 & 0 & \cdot\cdot & 1 & \cdot\cdot & 0 & \cdot\cdot & 1 \end{pmatrix}$$

is a $(I+J+1) \times IJ$ matrix of rank $I + J-1$.

As we have discussed in the case of multivariate one-way classification if $H_{\underline{\alpha}}$ is rejected it is necessary to test hypotheses of the form $H^*_{\underline{\alpha}}: \sum_i a_i \underline{\alpha}^i = \underline{\delta}$ where the a_i are real constants satisfying $\sum_i a_i = 0$ and $\underline{\delta}$ is a fixed vector. Let $\lambda^{-1} = \sum_i a_i^2/J$. Since $\lambda^{\frac{1}{2}} \sum_i a_i \underline{X}_{i.}$ has a p-variate normal distribution with mean $\lambda^{\frac{1}{2}} \sum_i a_i \underline{\alpha}^i$ and positive definite covariance matrix Σ and the random matrix C is distributed independently of $\lambda^{\frac{1}{2}} \sum_i a_i \underline{X}_{i.}$ as $W_p(\Sigma, (I-1)(J-1))$ we obtain from Theorem 2.14 that

$$(I-1)(J-1)\ \lambda\ (\sum_i a_i \underline{X}_{i.} - \underline{\delta})'\ C^{-1}(\sum_i a_i \underline{X}_{i.} - \underline{\delta})$$

is distributed as Hotelling's T^2-statistic with degree of freedom $(I-1)(J-1)$ and the non-centrality parameter $\lambda(\sum_i a_i \underline{\alpha}^i - \underline{\delta})'\Sigma^{-1}(\sum_i a_i \underline{\alpha}^i - \underline{\delta})'$. Thus the optimum invariant test of $H^*_{\underline{\alpha}}$ rejects $H^*_{\underline{\alpha}}$ (at level α) if

$$(I-1)(J-1)\ \lambda\ (\sum_i a_i \underline{x}_{i.} - \underline{\delta})'\ c^{-1}(\sum_i a_i \underline{x}_{i.} - \underline{\delta})$$

$$\geq \frac{(I-1)(J-1)p}{((I-1)(J-1)-p+1)} F_{1-\alpha,p,(I-1)(J-1)-p+1}.$$

Similarly if H_β is rejected the optimum invariant test of $H^*_{\underline{\beta}}$: $\sum_j a_j \underline{\beta}^j = \underline{\gamma}$, where the a_j are real constants satisfying $\sum_j a_j = 0$ and $\underline{\gamma}$ is a fixed vector, rejects $H^*_{\underline{\beta}}$ if

$$(I-1)(J-1)\ \ \lambda\ (\sum_j a_j \underline{X}_{.j} - \underline{\gamma})'\, C^{-1} (\sum_j a_j \underline{x}_{.j} - \underline{\gamma})$$

$$\geq \frac{(I-1)(J-1)p}{(I-1)(J-1)-p+1} F_{1-\alpha,p,(I-1)(J-1)-p+1}$$

where $\mu^{-1} = \sum_j a_j^2/I$. Note that $(I-1)(J-1)\ \ \mu\ (\sum_j a_j \underline{X}_{.j} - \underline{\gamma})'\, C^{-1} (\sum_j a_j \underline{X}_{.j} - \underline{\gamma})$ is distributed as Hotelling's T^2-statistic with noncentrality parameter

$$\mu = (\sum_j a_j\ \underline{\beta}^j - \underline{\gamma})'\, \Sigma^{-1} (\sum_j a_j \underline{\beta}^j - \underline{\gamma}) \ .$$

7.6.3. Multivariate Two-way Classifications with K Observations Per Cell: Suppose that we have a set of independently distributed normal p-vectors $\underline{X}_{ijk} = (X_{ijk1}, \ldots, X_{ijkp})'$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, $(K > 1)$ with

$$E(\underline{X}_{ijk}) = \underline{\mu} + \underline{\alpha}^i + \underline{\beta}^j + \underline{\gamma}^{ij}, \quad \mathrm{Cov}(\underline{X}_{ijk}) = \Sigma \text{ (positive definite)}$$

where

$$\underline{\mu} = (\mu_1, \ldots, \mu_p)', \quad \underline{\alpha}^i = (\alpha_{i1}, \ldots, \alpha_{ip})' \text{ with } \sum_i \underline{\alpha}^i = \underline{0},$$

$$\underline{\beta}^j = (\beta_{j1}, \ldots, \beta_{jp})' \text{ with } \sum_j \underline{\beta}^j = \underline{0} \quad \underline{\gamma}^{ij} = (\gamma_{ij1}, \ldots, \gamma_{ijp})' \text{ with}$$

$\sum_i \underline{\gamma}^{ij} = \sum_j \underline{\gamma}^{ij} = 0$. The $\underline{\gamma}^{ij}$ are interaction between the two factors of the two-way classification. In the notations of Section 7.1 H_α is a multivariate linear hypothesis with $r = I-1$, $s = IJ$, $N-s = IJ(K-1)$ and similarly for H_β. The hypothesis $H_{\underline{\gamma}}$ is a multivariate linear hypothesis with $r = (I-1)(J-1)$, $s = IJ$, $N-s = IJ(K-1)$. Under π_Ω the maximum likelihood estimate of $\underline{\mu}$, $\underline{\alpha}^i$, $\underline{\beta}^j$, $\underline{\gamma}^{ij}$, Σ are given by

$$\hat{\underline{\mu}} = \underline{x}_{...}, \quad \hat{\underline{\alpha}}^i = \underline{x}_{i..} - \underline{x}_{...}, \quad \hat{\underline{\beta}}^j = \underline{x}_{.j.} - \underline{x}_{...},$$

$$\hat{\gamma}^{ij} = (\underline{x}_{ij.}-\underline{x}_{i..}-\underline{x}_{.j.}+\underline{x}_{...}),$$

$$\hat{\Sigma} = (IJK)^{-1}\sum_i\sum_j\sum_k(\underline{x}_{ijk}-\underline{x}_{ij.})(\underline{x}_{ijk}-\underline{x}_{ij.})'$$

Under the hypothesis $H_{\underline{\alpha}}$ the maximum likelihood estimate of $\underline{\mu}$, $\underline{\beta}^j$, $\underline{\gamma}^{ij}$ and Σ are given by

$$\hat{\hat{\mu}} = \underline{x}_{...},\quad \hat{\hat{\underline{\beta}}}^j = \underline{x}_{.j.}-\underline{x}_{...},\quad \hat{\hat{\underline{\gamma}}}^{ij} = (\underline{x}_{ij.}-\underline{x}_{i..}-\underline{x}_{.j.}+\underline{x}_{...}),$$

$$\hat{\hat{\Sigma}} = (IJK)^{-1}\sum_i\sum_j\sum_k(\underline{x}_{ijk}-\underline{x}_{ij.}+\underline{x}_{i..}-\underline{x}_{...})(\underline{x}_{ijk}-\underline{x}_{ij.}+\underline{x}_{i..}-\underline{x}_{...})'$$

Let

$$G = \sum_i\sum_j\sum_k(\underline{X}_{ijk}-\underline{X}_{ij.})(\underline{X}_{ijk}-\underline{X}_{ij.})'$$

$$D = \sum_i\sum_j K(\underline{X}_{ij.}-\underline{X}_{i..}-X_{.j.}+\underline{X}_{...})(\underline{X}_{ij.}-\underline{X}_{i..}-\underline{X}_{.j.}+\underline{X}_{...})'$$

$$B = \sum_j IK(\underline{X}_{.j.}-\underline{X}_{...})(\underline{X}_{.j.}-\underline{X}_{...})'.$$

$$A = \sum_i JK(\underline{X}_{i..}-\underline{X}_{...})(\underline{X}_{i..}-\underline{X}_{...})'$$

and let g,d,b,a denote their sample values. The random matrices G,D,B,A are independently distributed Wishart matrices as $p \le I-1$, $p \le J-1$ by assumption. It may be easily checked that the likelihood ratio test of $H_{\underline{\alpha}}$ rejects $H_{\underline{\alpha}}$ if

$$u_1 = \frac{\det g}{\det(a+g)} \le k$$

where k is a constant depending on the level of significance α of the test under $H_{\underline{\alpha}}$ the statistic U_1(with values u_1) is distributed as $U_{p,I-1,IJ(K-1)}$.

Similarly the likelihood ratio rest of $H_{\underline{\beta}}$ rejects $H_{\underline{\beta}}$ whenever

$$u_2 = \frac{\det g}{\det(b+g)} \le k$$

where the constant k depends on the level of significance α of the test and under $H_{\underline{\beta}}$ the statistic U_2 (with values u_2) is distributed $U_{p,J-1,IJ(K-1)}$. The likelihood ratio test of $H_{\underline{\gamma}}$ rejects $H_{\underline{\gamma}}$ whenever

$$u_3 = \frac{\det g}{\det(d+g)} \le k$$

where the constant k depends on the level of significance α of the test and under $H_{\underline{\gamma}}$ the statistic U_3 (with values u_3) is distributed as $U_{p,(I-1)(J-1),IJ(K-1)}$. Other invariant tests of $H_{\underline{\alpha}}$, $H_{\underline{\beta}}$ and $H_{\underline{\gamma}}$ can be obtained following (7.14)-(7.17).

If $H_{\underline{\alpha}}$ is rejected one would be interested in testing hypothesis of the form $H^*_{\underline{\alpha}}: \sum_i a_i \underline{\alpha}^i = \underline{\delta}$ where the a_i are real satisfying $\sum_i a_i = 0$ and $\underline{\delta}$ is a fixed vector. The optimum invariant test for this problem is to reject $H^*_{\underline{\alpha}}$ if

$$\lambda^{-1}\, IJ(K-1)\Big(\sum_i a_i \underline{x}_{i..} - \underline{\delta}\Big)'\, g^{-1}\Big(\sum_i a_i \underline{x}_{i..} - \underline{\delta}\Big)$$

$$\ge \frac{IJ(K-1)p}{IJ(K-1)-p+1}\, F_{1-\alpha,p,IJ(K-1)-p+1}.$$

Similarly if H_β is rejected one would be interested to test hypothesis of the form $H^*_{\underline{\beta}}: \sum_i a_j \underline{\beta}^j = \underline{\gamma}$ where the a_i are reals satisfying $\sum_j a_j = 0$ and $\underline{\gamma}$ is a fixed vector. The optimum invariant test for this problem is to reject $H^*_{\underline{\beta}}$ if

$$\mu^{-1} IJ(K-1)\Big(\sum_j a_j \underline{x}_{.j.} - \underline{\gamma}\Big)'\, g^{-1}\Big(\sum_j a_j \underline{x}_{.j.} - \underline{\gamma}\Big)$$

$$\ge \frac{IJ(K-1)p}{IJ(K-1)-p+1}\, F_{1-\alpha,p,IJ(K-1)-p+1}.$$

If $H_{\underline{\gamma}}$ is rejected then one would not ordinarily explore it further, although theoretically one could use similar invariant test procedures as in the case of $H^*_{\underline{\alpha}}$ and $H^*_{\underline{\beta}}$.

If it is possible an effort should be made to collect equal number of observations per cell. The analysis of a two-way classification with unequal number of observations per cell is more complicated. Nevertheless, exact procedures used in the univariate case can be extended to the multivariate case in a straightforward manner.

Roy and Bose (1953) and Roy (1957) have studied the problem of construction of simultaneous confidence interval associated with the maximum characteristic root criterion of Roy for the means of multivariate analysis of variance problem.

Example 7.1. (Two-way classification) Table 7.1 contains Data on 12 different double crosses of barley which were raised during 1971-1972 in Hissar, India. Each observation vector has 6 components: (i) plant height in centimeters, (ii) length of ear head in centimeters, (iii) average number of ears per plant, (iv) average yield per plant in grams, (v) weight of one thousand such in grams, (vi) average ear weight in grams. The column indices run over different crosses of barley and the row indices run over four different locations.

Now $u = \det b/ \det(a+b) = .1449$,

$$-\{(I-1)(J-1) - \tfrac{1}{2}(p-I+2)\} \log u = 67.60985.$$

From Table 7.3 we obtain

One percent value of $\{-(I-1)(J-1)-\tfrac{1}{2}(p-I+2)\} \log U$ is $(1.02)(95.6257)$.

Five percent value of $\{-(I-1)(J-1)-\tfrac{1}{2}(p-I+2)\} \log U$ is $(1.02)(85.9649)$.

Thus our data show there are no significant differences between crosses.

.6.4. Multivariate Randomized Block Design: We have treated the univariate randomized block design in complete details in apter 5. For the multivariate case the layout is the same as the univariate one but each observation here is a p-dimensional vector $\underline{x}_{ij}=(x_{ij1}, \dots, x_{ijp})'$, $j = 1,\dots,\nu$, $i = 1,\dots,b$ (observation on p response variates) where b is the number of blocks each containing ν treatments. For the multivariate randomized block design the mathematical model is

$E(\underline{X}_{ij}) = \underline{\mu} + \underline{\alpha}^{i} + \underline{\beta}^{j}$, $\mathrm{Cov}(\underline{X}_{ij}) = \Sigma$ (positive definite)

where $\underline{x}_{ij}$ is the observed effect of the jth treatment in the ith block, $\underline{\mu} = (\mu_1,\dots,\mu_p)'$ is the grand mean, $\underline{\alpha}^{i} = (\alpha_{i1},\dots,\alpha_{ip})'$ is the effect of the ith block such that $\sum_i \underline{\alpha}^{i} = \underline{0}$, $\underline{\beta}^{j} = (\beta_{j1},\dots,\beta_{jp})'$ is the effect of jth treatment satisfying $\sum_j \underline{\beta}^{j} = \underline{0}$ and the $\underline{X}_{ij}$ are independently normally

Table 7.1. Double cross of barley

Locations	1	2	3	4	5	6	7	8	9	10	11	12
1	126.60 59.80 19.66 5.62 101.32 18.03	123.26 46.20 21.55 4.46 91.62 20.54	137.24 46.20 19.33 4.44 108.72 26.74	129.26 56.20 19.78 5.46 103.04 18.87	123.26 49.60 20.33 4.62 89.18 19.30	140.06 44.40 20.79 4.26 110.14 20.85	136.76 55.80 20.65 5.04 91.78 18.21	112.86 51.40 20.36 5.08 105.32 20.72	129.64 48.20 19.09 4.22 109.28 25.90	121.40 52.00 20.28 4.96 90.22 18.19	121.26 43.40 20.60 4.18 80.26 19.20	134. 46.80 19.57 3.96 104.36 26.35
2	133.04 39.60 19.95 3.34 77.08 23.08	133.96 52.20 21.24 4.40 82.62 18.78	130.50 52.00 20.65 4.64 85.46 18.42	126.26 39.80 21.54 3.58 76.36 22.33	127.58 51.20 20.43 4.52 78.74 17.42	119.36 42.20 19.94 4.06 69.64 17.15	128.54 45.60 20.31 3.92 97.40 24.85	118.42 44.40 20.54 4.28 78.74 18.40	128.04 49.20 19.78 4.56 86.28 18.92	122.48 40.80 20.52 3.84 95.32 24.82	133.72 52.60 20.24 4.96 84.78 17.09	121.98 47.80 20.15 4.12 68.96 16.74
3	113.90 37.60 20.99 3.22 91.96 28.56	125.42 39.60 19.40 3.56 72.18 20.27	127.96 42.60 20.06 3.86 80.32 20.82	115.82 28.20 21.64 2.46 68.16 27.70	133.74 40.60 19.01 3.78 74.02 19.58	121.26 52.80 19.37 4.48 92.66 20.68	107.28 39.40 19.10 3.04 85.62 28.16	137.08 39.20 20.73 3.46 72.28 20.87	116.06 48.00 19.54 4.36 96.76 22.19	118.32 36.20 20.32 3.28 96.16 29.32	129.56 44.40 20.34 4.18 82.74 19.79	127.08 53.80 20.02 4.96 114.14 23.01
4	121.52 52.40 20.03 5.12 92.46 18.06	128.06 31.60 20.65 2.78 77.68 27.94	129.24 48.00 19.69 4.36 90.48 20.75	125.10 61.60 20.00 5.34 99.66 18.66	133.82 38.80 19.23 3.12 82.96 26.42	130.78 49.20 20.27 4.28 96.12 22.46	132.56 52.40 19.44 4.62 77.64 16.81	127.96 41.60 19.52 3.72 93.66 25.18	137.12 44.80 20.44 3.94 92.44 23.46	127.64 55.80 20.50 4.94 88.42 17.90	127.92 38.60 20.46 3.56 90.48 25.42	132.28 47.60 20.4 4.4 98.2 21.

The matrix a, b are given by

$$a = \begin{pmatrix} 772958.19120 & & & & & \\ 281956.78700 & 103035.87000 & & & & \\ 122836.93360 & 44814.78400 & 19540.58862 & & & \\ 25553.52290 & 9335.66200 & 4062.63695 & 846.32530 & & \\ 542196.30440 & 198199.66200 & 96191.13345 & 17964.13050 & 38170.87290 & \\ 131499.07676 & 48016.45650 & 20914.98705 & 4351.16635 & 92390.13750 & 22418.60407 \end{pmatrix}$$

$$b = \begin{pmatrix} 774437.42900 & & & & & \\ 282055.10200 & 104619.95500 & & & & \\ 122862.63635 & 44800.21258 & 19548.35585 & & & \\ 25560.87220 & 9488.29383 & 4060.99354 & 861.83688 & & \\ 542398.55150 & 198624.29583 & 86200.81826 & 18005.83032 & 383842.35742 & \\ 131452.59180 & 47300.86142 & 20919.62470 & 4277.39994 & 92690.46031 & 22903.06659 \end{pmatrix}$$

distributed. The analysis of this design with fixed effect model can be done in the same way as the analysis of a multivariate two-way classification with one observation per cell without any block-treatment interactions.

7.6.5. Multivariate Latin Squares Design: We have described in details the different aspects of univariate latin square designs in Chapter 5. The layout for the multivariate case is the same as the univariate one but the observation is a p-dimensional vector (observations on p responses) $\underline{x}_{ijk} = (x_{ijk1}, \dots, x_{ijkp})'$, $i, j, k = 1, \dots, \nu$, where $\underline{x}_{ijk}$ is the observation vector on the kth treatment in the ith row and the jth column. The fixed effect mathematical model of the multivariate latin square design is given by

$$E(\underline{X}_{ijk}) = \underline{\mu} + \underline{\alpha}^{i} + \underline{\beta}^{j} + \underline{t}^{k}, \quad \mathrm{Cov}(\underline{X}_{ijk}) = \Sigma \quad \text{(positive definite)}$$

where the triplets (i, j, k) take on only the set S of ν^2 values indicated by the particular latin square selected for the experiment, $\underline{\mu} = (\mu_1, \dots, \mu_p)'$ is the grand mean, $\underline{\alpha}^{i} = (\alpha_{i1}, \dots, \alpha_{ip})'$ is the ith row effect such that $\sum_i \underline{\alpha}^{i} = \underline{0}$, $\underline{\beta}^{j} = (\beta_{j1}, \dots, \beta_{jp})'$ is the jth column effect such that $\sum_j \underline{\beta}^{j} = \underline{0}$, and $\underline{t}^{k} = (t_{k1}, \dots, t_{kp})'$ is the kth treatment effect such that $\sum_k \underline{t}^{k} = \underline{0}$.

In this design the hypotheses of interest are

$H_{\underline{\alpha}}$: all $\underline{\alpha}^{i} = 0$, $H_{\underline{\beta}}$: all $\underline{\beta}^{j} = 0$, $H_{\underline{t}}$: all $\underline{t}^{k} = 0$.

In the notation of 7.1 each hypothesis is a multivariate linear hypothesis with $r = \nu-1$, $s = 1+3\nu-3 = 3\nu-2$, $N-s=\nu^2-3\nu+2 = (\nu-1)(\nu-2)$. Let

$$\underline{x}_{...} = \sum_{(i,j,k)\in S} \underline{x}_{ijk}, \quad \underline{x}_{i..} = \sum_{(j,k)\in S} \underline{x}_{ijk}, \quad \underline{x}_{.j.} = \sum_{(i,k)\in S} \underline{x}_{ijk}, \quad \underline{x}_{..k} = \sum_{(i,j)\in S} \underline{x}_{ijk}$$

Under π_{Ω} the maximum likelihood estimate of $\underline{\mu}$, $\underline{\alpha}^{i}$, $\underline{\beta}^{j}$, $\underline{t}^{k}$, Σ are given by

$$\hat{\underline{\mu}} = \underline{x}_{...},\ \hat{\underline{\alpha}}^{i} = \underline{x}_{i..}-\underline{x}_{...},\ \hat{\underline{\beta}}^{j} = \underline{x}_{.j.}-\underline{x}_{...},\ \hat{\underline{t}}^{k}=\underline{x}_{..k}-\underline{x}_{...},$$

$$\nu^2\hat{\Sigma} = \sum_{(i,j,k)\varepsilon S} \{(\underline{x}_{ijk}-\underline{x}_{i..}-\underline{x}_{.j.}-\underline{x}_{..k}+2\underline{x}_{...}) \times (\underline{x}_{ijk}-\underline{x}_{i..}-\underline{x}_{.j.}-\underline{x}_{..k}+2\underline{x}_{...})'\} = g \quad \text{(say)}. \tag{7.76}$$

The random matrix G (with values g) is distributed as $W_p(\Sigma, (\nu-1)(\nu-2))$ provided that $(\nu-1)(\nu-2) \geq p$. Under the hypothesis $H_{\underline{t}}$ the maximum likelihood estimates of $\underline{\mu}, \underline{\alpha}^{i}, \underline{\beta}^{j}, \Sigma$ are given by

$$\hat{\underline{\mu}} = \underline{x}_{...},\ \hat{\underline{\alpha}}^{i} = \underline{x}_{i..}-\underline{x}_{...},\ \hat{\underline{\beta}}^{j}=\underline{x}_{.j.}-\underline{x}_{...},$$

$$\begin{aligned}\nu^2\hat{\hat{\Sigma}} &= \sum_{(i,j,k)\varepsilon S} (\underline{x}_{ijk}-\underline{x}_{i..}-\underline{x}_{.j.}+\underline{x}_{...})(\underline{x}_{ijk}-\underline{x}_{i..}-\underline{x}_{.j.}+\underline{x}_{...})' \\ &= \nu^2 + \sum_{k} (\underline{x}_{..k}-\underline{x}_{...})(\underline{x}_{..k}-\underline{x}_{...})' \\ &= g+c \quad \text{(say)}.\end{aligned} \tag{7.77}$$

The random matrices G, C (with values c) are independent and under $H_{\underline{t}}$ C is distributed as $W_p(\Sigma, \nu-1)$ provided that $\nu-1 \geq p$. From (7.76)-(7.77) the likelihood ratio test of $H_{\underline{t}}$ rejects $H_{\underline{t}}$ whenever

$$u_1 = \frac{\det g}{\det(g+c)} \leq k$$

where the constant k depends on the level of significance α of the test and under $H_{\underline{t}}$ the statistic U_1 (with values u_1) is distributed as $U_{p,\nu-1,(\nu-1)(\nu-2)}$.

Similarly the likelihood ratio test of $H_{\underline{\alpha}}$ rejects $H_{\underline{\alpha}}$ whenever

$$u_2 = \frac{\det g}{\det(a+g)} \leq k$$

where $a = \sum_i \nu\,(\underline{x}_{i..}-\underline{x}_{...})(\underline{x}_{i..}-\underline{x}_{...})'$ and k is a constant depending on the level of significance α of the test and under $H_{\underline{\alpha}}$ the statistic U_2 (with values u_2) is distributed as $U_{p,\nu-1,(\nu-1)(\nu-2)}$.

The likelihood ratio test of the hypothesis $H_{\underline{\beta}}$ rejects $H_{\underline{\beta}}$ whenever

$$u_3 = \frac{\det g}{\det(b+g)} \le k$$

where $b = \sum_j \nu(\underline{x}_{.j.} - \underline{x}_{...})(\underline{x}_{.j.} - \underline{x}_{...})'$ and k is a constant depending on the level of significance α of the test and under $H_{\underline{\alpha}}$ the statistic U_3 (with values u_3) is distributed as $U_{p,\nu-1,(\nu-1)(\nu-2)}$.

Other invariant tests of $H_{\underline{\alpha}}$, $H_{\underline{\beta}}$ and $H_{\underline{t}}$ can be obtained following (7.14)-(7.17).

If any of the hypotheses, say, $H_{\underline{t}}$ is rejected then we need to test hypothesis of the form $H^*_{\underline{t}}: \sum_k c_k \underline{t}^k = \underline{\sigma}$ where the c_k are real constants such that $\sum_k c_k = 0$ and $\underline{\sigma}$ is a fixed vector. The optimum invariant test of $H^*_{\underline{t}}$ rejects $H^*_{\underline{t}}$ whenever

$$(\nu-1)(\nu-2)\ \lambda\ (\sum_k c_k\, \underline{x}_{..k} - \underline{\sigma})'\ g^{-1} (\sum_k c_k \underline{x}_{..k} - \underline{\sigma})$$

$$\ge \frac{(\nu-1)(\nu-2)}{(\nu-1)(\nu-2)-p+1}\ F_{1-\alpha,p,(\nu-1)(\nu-2)-p+1},$$

where $\lambda^{-1} = \sum_k \frac{c_k^2}{\nu}$. Note that

$$(\nu-1)(\nu-2)\lambda(\sum_k c_k\, \underline{X}_{...k} - \underline{\sigma})'\ G^{-1}\ (\sum_k c_k\, \underline{X}_{..k} - \underline{\sigma})$$

is distributed as Hotelling's T^2-statistic with degrees of freedom $(\nu-1)(\nu-2)$ and noncentrality parameter

$$\lambda(\sum_k c_k \underline{t}^k - \underline{\sigma})'\ \Sigma^{-1}\ (\sum_k c_k \underline{t}^k - \underline{\sigma}).$$

The rejection of $H_{\underline{\alpha}}$ or $H_{\underline{\beta}}$ or both should be similarly treated.

7.7. MULTIVARIATE ANALYSIS OF COVARIANCE

We have treated the analysis of covariance of univariate one-way and two-way classifications in Chapter 4 and those of different designs in Chapter 5. For the multivariate analysis of covariance the concepts are the same but the observations and parameters are vectors. As in the univariate case there is a multivariate analysis of covariance model for every multivariate analyses of variance model. For example the analysis of covariance model for the two-way classification without interaction can be written as

$$E(\underline{X}_{ij}) = \underline{\mu} + \underline{\alpha}^{i} + \underline{\beta}^{j} + \Delta\underline{u}_{ij}, \quad i = 1,\ldots,I, \ j = 1,\ldots,J,$$

$$\mathrm{Cov}(\underline{X}_{ij}) = \Sigma \text{ (positive definite)},$$

or equivalently

$$E(X) = \beta z + \Delta u = (\beta,\Delta)\binom{z}{u},$$

$$\mathrm{Cov}(\underline{X}_{ij}) = \Sigma$$

where β, z are as defined in (7.75) and $\underline{u}_{ij} = (u_{ij1},\ldots,u_{ijk})'$ is a vector of k concomitant variates

$$u = (\underline{u}_{11},\ldots,\underline{u}_{1J},\ldots,\underline{u}_{I1},\ldots,\underline{u}_{IJ})$$

is a $k \times (IJ)$ matrix and Δ is a $p \times k$ matrix of unknown coefficients which are assumed to be the same for all (i,j). Under this model the design matrix is $\binom{z}{u}$ and (β,Δ) is the parameter matrix. Thus, as in the case of univariate analysis of covariance, the multivariate analysis of covariance can be treated as the multivariate regression model as discussed in Section 7.5. The problem of estimation and testing hypotheses on (β,Δ) can be resolved from the methodology developed in Section 7.5.

7.8. CASE OF MULTINOMIAL AND COMPLEX MULTIVARIATE NORMAL DISTRIBUTIONS

Multivariate analysis of variance of one-way, two-way and higher-way classifications with discrete data following multinomial distributions has been treated by Roy (1957), Lehmann (1959), and Rao (1965).

We shall give some recent developments regarding the complex multivariate general linear hypothesis which is defined for the complex multivariate normal distributions in the same way as the multivariate linear hypothesis for the multivariate normal distribution. The test statistic depends, as in the real case, on the characteristic roots of two independent complex Wishart matrices. The distribution of statistics based on the characteristic roots of complex Wishart type matrices is also useful in multiple time series analysis (Hannan 1970). The joint noncentral distribution of the characteristic roots of complex Wishart matrices associated with the complex multivariate general linear hypothesis model were given explicitly by James (1964) in terms of Zonal polynomials and by Khatri (1964a) in the form of integrals. In the case of central complex Wishart matrices and complex matrices connected with the complex multivariate general linear hypothesis, the distribution of extreme characteristic roots were obtained by Pillai and Young (1970) and Pillai and Jouris (1972). Khatri (1946b, 1970) obtained the noncentral distribution of the individual characteristic roots of the complex matrices associated with the complex general linear hypothesis and that of traces.

7.9. STATISTICAL TABLES ON WILKS' CRITERION $U=U_{p,r,N-s}$

In this section we shall reproduce, with the kind permission of the Editor, Biometrika, the tables of Schatzoff (1966) and of Pillai and Gupta (1969) for the chisquare adjustment factor

$$C = \frac{[\text{ percentile for } \{-(N-s-\tfrac{1}{2}(p-r+1))\log_e U\}]}{\text{percentile for } \chi^2_{pr}}$$

for converting the upper percentile points of χ^2_{pr} to get the exact percentile points of $\{-(N-s-\tfrac{1}{2}(p-r+1)\log_e U\}$ (lower percentile point of U) for different values of $p, r, M = N-s-p+1$. In other words to obtain the required percentile point $-(N-s-\tfrac{1}{2}(p-r+1))\log_e U$ we multiply the corresponding upper percentile point of χ^2_{pr} by the tabulated value of the adjustment factor C. Table 7.2 is due to Schatzoff (1966) and Table 7.3 is due to Pillai and Gupta (1969).

Table 7.2

Tables of chi-square adjustments to Wilks's criterion U. Factor C for lower percentiles of U(upper percentiles of χ^2_{pr})

$p = 3, r = 4$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·379	1·422	1·463	1·514	1·550
2	1·159	1·174	1·188	1·207	1·220
3	1·091	1·099	1·107	1·116	1·123
4	1·060	1·065	1·070	1·076	1·080
5	1·043	1·046	1·050	1·054	1·057
6	1·032	1·035	1·037	1·040	1·042
7	1·025	1·027	1·029	1·031	1·033
8	1·020	1·022	1·023	1·025	1·026
9	1·017	1·018	1·019	1·021	1·022
10	1·014	1·015	1·016	1·017	1·018
12	1·010	1·011	1·012	1·012	1·013
14	1·008	1·008	1·009	1·010	1·010
16	1·006	1·007	1·007	1·007	1·008
18	1·005	1·005	1·006	1·006	1·006
20	1·004	1·004	1·005	1·005	1·005
24	1·003	1·003	1·003	1·004	1·004
30	1·002	1·002	1·002	1·002	1·002
40	1·001	1·001	1·001	1·001	1·001
60	1·000	1·001	1·001	1·001	1·001
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	18·5494	21·0261	23·3367	26·2170	28·2995

$p = 4, r = 4$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·405	1·451	1·494	1·550	1·589
2	1·178	1·194	1·209	1·229	1·243
3	1·105	1·114	1·122	1·132	1·139
4	1·071	1·076	1·081	1·088	1·092
5	1·051	1·055	1·058	1·063	1·066
6	1·039	1·042	1·044	1·048	1·050
7	1·031	1·033	1·035	1·037	1·039
8	1·025	1·027	1·028	1·030	1·032
9	1·020	1·022	1·023	1·025	1·026
10	1·017	1·018	1·019	1·021	1·022
12	1·013	1·014	1·014	1·015	1·016
14	1·010	1·010	1·011	1·012	1·012
16	1·008	1·008	1·009	1·009	1·010
18	1·006	1·007	1·007	1·008	1·008
20	1·005	1·006	1·006	1·006	1·007
24	1·004	1·004	1·004	1·004	1·005
30	1·002	1·003	1·003	1·003	1·003
40	1·001	1·002	1·002	1·002	1·002
60	1·001	1·001	1·001	1·001	1·001
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	23·5418	26·2962	28·8454	31·9999	34·2672

$p = 5, r = 4$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·435	1·483	1·530	1·589	1·632
2	1·199	1·216	1·233	1·253	1·269
3	1·121	1·130	1·139	1·150	1·158
4	1·083	1·089	1·094	1·101	1·106
5	1·061	1·065	1·069	1·074	1·077
6	1·047	1·050	1·053	1·056	1·059
7	1·037	1·040	1·042	1·044	1·046
8	1·030	1·032	1·034	1·036	1·038
9	1·025	1·027	1·028	1·030	1·031
10	1·021	1·023	1·024	1·025	1·026
12	1·016	1·017	1·018	1·019	1·020
14	1·012	1·013	1·014	1·014	1·015
16	1·010	1·010	1·011	1·012	1·012
18	1·008	1·008	1·009	1·009	1·010
20	1·007	1·007	1·007	1·008	1·008
24	1·005	1·005	1·005	1·006	1·006
30	1·003	1·003	1·004	1·004	1·004
40	1·002	1·002	1·002	1·002	1·002
60	1·001	1·001	1·001	1·001	1·001
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	28·4120	31·4104	34·1696	37·5662	39·9968

$p = 6, r = 4$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·466	1·517	1·566	1·628	1·674
2	1·222	1·240	1·257	1·279	1·295
3	1·138	1·148	1·157	1·168	1·177
4	1·096	1·102	1·108	1·115	1·121
5	1·071	1·076	1·080	1·085	1·089
6	1·055	1·059	1·062	1·066	1·068
7	1·044	1·047	1·049	1·052	1·055
8	1·036	1·038	1·040	1·043	1·045
9	1·030	1·032	1·034	1·036	1·037
10	1·026	1·027	1·029	1·030	1·032
12	1·019	1·020	1·021	1·023	1·024
14	1·015	1·016	1·017	1·018	1·018
16	1·012	1·013	1·013	1·014	1·015
18	1·010	1·010	1·011	1·012	1·012
20	1·008	1·009	1·009	1·010	1·010
24	1·006	1·006	1·007	1·007	1·007
30	1·004	1·004	1·004	1·005	1·005
40	1·002	1·002	1·003	1·003	1·003
60	1·001	1·001	1·001	1·001	1·001
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	33·1963	36·4151	39·3041	42·9798	45·5585

Table 7.2 (cont.)

$p = 7$, r = 4

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·497	1·550	1·601	1·667	1·715
2	1·244	1·263	1·281	1·305	1·322
3	1·155	1·165	1·175	1·188	1·197
4	1·109	1·116	1·122	1·130	1·136
5	1·082	1·087	1·092	1·097	1·101
6	1·064	1·068	1·071	1·076	1·079
7	1·052	1·055	1·057	1·061	1·063
8	1·043	1·045	1·047	1·050	1·052
9	1·036	1·038	1·040	1·042	1·044
10	1·031	1·032	1·034	1·036	1·037
12	1·023	1·024	1·026	1·027	1·028
14	1·018	1·019	1·020	1·021	1·022
16	1·015	1·015	1·016	1·017	1·017
18	1·012	1·013	1·013	1·014	1·014
20	1·010	1·011	1·011	1·012	1·012
24	1·007	1·008	1·008	1·008	1·009
30	1·005	1·005	1·005	1·006	1·006
40	1·003	1·003	1·003	1·003	1·004
60	1·001	1·001	1·002	1·002	1·002
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	37·9159	41·3372	44·4607	48·2782	50·9933

$p = 8$, r = 4

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·528	1·583	1·636	1·704	1·754
2	1·266	1·286	1·305	1·330	1·348
3	1·172	1·183	1·193	1·207	1·216
4	1·123	1·130	1·137	1·146	1·152
5	1·093	1·099	1·103	1·109	1·114
6	1·074	1·078	1·081	1·086	1·089
7	1·060	1·063	1·066	1·070	1·072
8	1·050	1·052	1·055	1·058	1·060
9	1·042	1·044	1·046	1·048	1·050
10	1·036	1·038	1·039	1·041	1·043
12	1·027	1·029	1·030	1·031	1·033
14	1·021	1·023	1·023	1·025	1·026
16	1·017	1·018	1·019	1·020	1·021
18	1·014	1·015	1·016	1·016	1·017
20	1·012	1·013	1·013	1·014	1·014
24	1·009	1·009	1·010	1·010	1·010
30	1·006	1·006	1·007	1·007	1·007
40	1·003	1·004	1·004	1·004	1·004
60	1·002	1·002	1·002	1·002	1·002
120	1·000	1·000	1·000	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	42·5847	46·1943	49·4804	53·4858	56·3281

$p = 9$, r = 4

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·557	1·614	1·669	1·740	1·792
2	1·288	1·309	1·329	1·355	1·373
3	1·189	1·201	1·212	1·223	1·236
4	1·137	1·144	1·152	1·161	1·167
5	1·105	1·110	1·115	1·122	1·127
6	1·083	1·088	1·091	1·096	1·100
7	1·068	1·071	1·075	1·078	1·081
8	1·057	1·060	1·062	1·065	1·068
9	1·048	1·050	1·053	1·055	1·057
10	1·041	1·043	1·045	1·047	1·049
12	1·032	1·033	1·034	1·036	1·037
14	1·025	1·026	1·027	1·029	1·029
16	1·020	1·021	1·022	1·023	1·024
18	1·017	1·018	1·018	1·019	1·020
20	1·014	1·015	1·015	1·016	1·017
24	1·010	1·011	1·011	1·012	1·012
30	1·007	1·007	1·008	1·008	1·008
40	1·004	1·004	1·005	1·005	1·005
60	1·002	1·002	1·002	1·002	1·002
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	47·2122	50·9985	54·4373	58·6192	61·5812

$p = 10$, r = 4

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·585	1·644	1·701	1·774	1·828
2	1·309	1·331	1·352	1·379	1·398
3	1·206	1·218	1·230	1·244	1·255
4	1·150	1·159	1·166	1·176	1·183
5	1·116	1·122	1·128	1·134	1·139
6	1·093	1·097	1·102	1·107	1·111
7	1·076	1·080	1·083	1·088	1·090
8	1·064	1·067	1·070	1·073	1·076
9	1·054	1·057	1·059	1·062	1·064
10	1·047	1·049	1·051	1·054	1·055
12	1·036	1·038	1·039	1·041	1·042
14	1·029	1·030	1·031	1·033	1·034
16	1·023	1·024	1·025	1·026	1·027
18	1·019	1·020	1·021	1·022	1·023
20	1·016	1·017	1·018	1·019	1·019
24	1·012	1·013	1·013	1·014	1·014
30	1·008	1·009	1·009	1·009	1·010
40	1·005	1·005	1·005	1·006	1·006
60	1·002	1·003	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	51·8050	55·7585	59·3417	63·6907	66·7659

Table 7.2 (cont.)

$p = 3, r = 6$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·482	1·535	1·586	1·649	1·694
2	1·222	1·241	1·259	1·282	1·298
3	1·135	1·145	1·155	1·167	1·176
4	1·092	1·099	1·105	1·113	1·119
5	1·068	1·072	1·077	1·082	1·086
6	1·052	1·056	1·059	1·063	1·066
7	1·041	1·044	1·047	1·050	1·052
8	1·034	1·036	1·038	1·041	1·042
9	1·028	1·030	1·032	1·034	1·035
10	1·024	1·025	1·027	1·028	1·030
12	1·018	1·019	1·020	1·021	1·022
14	1·014	1·014	1·015	1·016	1·017
16	1·011	1·012	1·012	1·013	1·014
18	1·009	1·009	1·010	1·011	1·011
20	1·007	1·008	1·008	1·009	1·009
24	1·005	1·006	1·006	1·006	1·007
30	1·004	1·004	1·004	1·004	1·004
40	1·002	1·002	1·002	1·002	1·003
60	1·001	1·001	1·001	1·001	1·001
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	25·9894	28·8693	31·5264	34·8053	37·1564

$p = 4, r = 6$

The values for C are obtained by entering the table for p = 6, r = 4, with M = N−s−3

$p = 5, r = 6$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·465	1·514	1·563	1·625	1·671
2	1·228	1·245	1·262	1·284	1·300
3	1·144	1·154	1·163	1·175	1·183
4	1·102	1·108	1·114	1·121	1·127
5	1·077	1·081	1·085	1·090	1·094
6	1·060	1·063	1·066	1·070	1·073
7	1·048	1·051	1·053	1·056	1·059
8	1·040	1·042	1·044	1·046	1·048
9	1·034	1·035	1·037	1·039	1·040
10	1·029	1·030	1·031	1·033	1·034
12	1·022	1·023	1·024	1·025	1·026
14	1·017	1·018	1·019	1·019	1·020
16	1·014	1·014	1·015	1·016	1·016
18	1·011	1·012	1·012	1·013	1·013
20	1·009	1·010	1·010	1·011	1·011
24	1·007	1·007	1·007	1·008	1·008
30	1·005	1·005	1·005	1·005	1·005
40	1·003	1·003	1·003	1·003	1·003
60	1·001	1·001	1·001	1·001	1·002
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	40·2560	43·7730	46·9792	50·8922	53·6720

$p = 6, r = 6$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·471	1·520	1·568	1·631	1·677
2	1·237	1·255	1·272	1·294	1·310
3	1·153	1·163	1·172	1·183	1·192
4	1·109	1·116	1·122	1·129	1·134
5	1·083	1·088	1·092	1·097	1·101
6	1·066	1·069	1·072	1·076	1·079
7	1·053	1·056	1·058	1·061	1·064
8	1·044	1·046	1·048	1·051	1·053
9	1·037	1·039	1·041	1·043	1·044
10	1·032	1·034	1·035	1·037	1·038
12	1·024	1·025	1·026	1·028	1·029
14	1·019	1·020	1·021	1·022	1·022
16	1·015	1·016	1·017	1·018	1·018
18	1·013	1·013	1·014	1·014	1·015
20	1·011	1·011	1·012	1·012	1·013
24	1·008	1·008	1·009	1·009	1·009
30	1·005	1·006	1·006	1·006	1·006
40	1·003	1·003	1·003	1·004	1·004
60	1·002	1·002	1·002	1·002	1·002
120	1·000	1·000	1·000	1·000	1·000
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	47·2122	50·0985	54·4373	58·6192	61·5812

Table 7.2. (cont.)

$p = 7, r = 6$

α M	0·100	0·050	0·025	0·010	0·005
1	1·481	1·530	1·579	1·642	1·688
2	1·249	1·266	1·284	1·306	1·322
3	1·163	1·173	1·182	1·194	1·203
4	1·118	1·124	1·131	1·138	1·144
5	1·090	1·095	1·099	1·105	1·109
6	1·072	1·075	1·079	1·083	1·086
7	1·059	1·062	1·064	1·067	1·070
8	1·049	1·051	1·053	1·056	1·058
9	1·042	1·043	1·045	1·047	1·049
10	1·036	1·037	1·039	1·041	1·042
12	1·027	1·029	1·030	1·031	1·032
14	1·022	1·023	1·023	1·024	1·025
16	1·018	1·018	1·019	1·020	1·020
18	1·014	1·015	1·016	1·016	1·017
20	1·012	1·013	1·013	1·014	1·014
24	1·009	1·009	1·010	1·010	1·010
30	1·006	1·006	1·007	1·007	1·007
40	1·004	1·004	1·004	1·004	1·004
60	1·002	1·002	1·002	1·002	1·002
120	1·000	1·000	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	54·0902	58·1240	61·7768	66·2062	69·3360

$p = 8, r = 6$

α M	0·100	0·050	0·025	0·010	0·005
1	1·494	1·543	1·592	1·656	1·703
2	1·261	1·279	1·297	1·319	1·336
3	1·174	1·184	1·194	1·205	1·214
4	1·127	1·134	1·140	1·148	1·153
5	1·098	1·103	1·108	1·113	1·117
6	1·079	1·082	1·086	1·090	1·093
7	1·065	1·068	1·070	1·074	1·076
8	1·054	1·057	1·059	1·062	1·063
9	1·046	1·048	1·050	1·052	1·054
10	1·040	1·042	1·043	1·045	1·046
12	1·031	1·032	1·033	1·035	1·036
14	1·024	1·025	1·026	1·027	1·028
16	1·020	1·021	1·021	1·022	1·022
18	1·016	1·017	1·018	1·018	1·019
20	1·014	1·014	1·015	1·015	1·016
24	1·010	1·011	1·011	1·011	1·012
30	1·007	1·007	1·008	1·008	1·008
40	1·004	1·004	1·005	1·005	1·005
60	1·002	1·002	1·002	1·002	1·002
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	60·9066	65·1708	69·0226	73·6826	76·9688

$p = 9, r = 6$

α M	0·100	0·050	0·025	0·010	0·005
1	1·508	1·558	1·607	1·671	1·719
2	1·275	1·293	1·311	1·333	1·350
3	1·185	1·196	1·205	1·218	1·227
4	1·137	1·144	1·150	1·158	1·164
5	1·107	1·112	1·116	1·122	1·126
6	1·086	1·090	1·093	1·098	1·101
7	1·071	1·074	1·077	1·080	1·083
8	1·060	1·062	1·065	1·067	1·069
9	1·051	1·053	1·055	1·058	1·059
10	1·044	1·046	1·048	1·050	1·051
12	1·034	1·035	1·036	1·037	1·037
14	1·027	1·028	1·029	1·031	1·031
16	1·022	1·023	1·024	1·024	1·025
18	1·019	1·019	1·020	1·020	1·021
20	1·016	1·016	1·017	1·017	1·018
24	1·012	1·012	1·013	1·013	1·013
30	1·008	1·008	1·009	1·009	1·009
40	1·005	1·005	1·005	1·005	1·006
60	1·002	1·002	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·002
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	67·6728	72·1532	76·1921	81·0688	84·5016

$p = 10, r = 6$

α M	0·100	0·050	0·025	0·010	0·005
1	1·523	1·573	1·623	1·687	1·736
2	1·288	1·307	1·325	1·348	1·365
3	1·197	1·208	1·218	1·230	1·239
4	1·147	1·154	1·161	1·169	1·175
5	1·115	1·120	1·125	1·131	1·135
6	1·093	1·097	1·101	1·106	1·109
7	1·078	1·081	1·084	1·087	1·090
8	1·066	1·068	1·071	1·074	1·076
9	1·056	1·059	1·061	1·063	1·065
10	1·049	1·051	1·053	1·055	1·056
12	1·038	1·040	1·041	1·042	1·043
14	1·031	1·032	1·033	1·034	1·034
16	1·025	1·026	1·026	1·027	1·028
18	1·021	1·021	1·022	1·023	1·023
20	1·018	1·018	1·019	1·019	1·020
24	1·013	1·014	1·014	1·015	1·015
30	1·009	1·010	1·010	1·010	1·010
40	1·006	1·006	1·006	1·006	1·006
60	1·003	1·003	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	74·3970	79·0819	83·2976	88·3794	91·9517

Table 7.2. (cont.)

$p = 3, r = 8$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·572	1·632	1·690	1·763	1·816
2	1·280	1·302	1·324	1·350	1·370
3	1·177	1·190	1·201	1·216	1·227
4	1·125	1·133	1·141	1·150	1·157
5	1·094	1·100	1·105	1·112	1·117
6	1·073	1·078	1·082	1·087	1·091
7	1·059	1·063	1·066	1·070	1·073
8	1·049	1·052	1·054	1·058	1·060
9	1·041	1·043	1·046	1·048	1·050
10	1·035	1·037	1·039	1·041	1·043
12	1·026	1·028	1·029	1·031	1·032
14	1·021	1·022	1·023	1·024	1·025
16	1·017	1·018	1·018	1·019	1·020
18	1·014	1·014	1·015	1·016	1·017
20	1·011	1·012	1·013	1·013	1·014
24	1·008	1·009	1·009	1·010	1·010
30	1·006	1·006	1·006	1·007	1·007
40	1·003	1·004	1·004	1·004	1·004
60	1·002	1·002	1·002	1·002	1·002
120	1·000	1·000	1·000	1·000	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	33·1963	36·4151	39·3641	42·9798	45·5585

$p = 5, r = 8$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·505	1·556	1·607	1·672	1·721
2	1·261	1·280	1·298	1·321	1·338
3	1·171	1·182	1·192	1·204	1·213
4	1·124	1·131	1·137	1·145	1·151
5	1·095	1·100	1·105	1·110	1·114
6	1·076	1·079	1·083	1·087	1·090
7	1·062	1·065	1·068	1·071	1·073
8	1·052	1·054	1·056	1·059	1·061
9	1·044	1·046	1·048	1·050	1·052
10	1·038	1·039	1·041	1·043	1·044
12	1·029	1·030	1·031	1·033	1·034
14	1·023	1·024	1·025	1·026	1·027
16	1·018	1·019	1·020	1·021	1·022
18	1·015	1·016	1·017	1·017	1·018
20	1·013	1·013	1·014	1·015	1·015
24	1·009	1·010	1·010	1·011	1·011
30	1·006	1·007	1·007	1·007	1·008
40	1·004	1·004	1·004	1·004	1·005
60	1·002	1·002	1·002	1·002	1·002
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	51·8050	55·7585	59·3417	63·6907	66·7659

$p = 7, r = 8$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·490	1·538	1·586	1·648	1·694
2	1·265	1·282	1·299	1·321	1·337
3	1·179	1·189	1·198	1·210	1·218
4	1·132	1·139	1·145	1·152	1·158
5	1·103	1·108	1·112	1·117	1·121
6	1·083	1·086	1·090	1·094	1·097
7	1·068	1·071	1·074	1·077	1·080
8	1·058	1·060	1·062	1·065	1·067
9	1·049	1·051	1·053	1·055	1·057
10	1·043	1·044	1·046	1·048	1·049
12	1·033	1·034	1·035	1·037	1·038
14	1·026	1·027	1·028	1·028	1·029
16	1·021	1·022	1·023	1·023	1·024
18	1·018	1·018	1·019	1·020	1·020
20	1·015	1·016	1·016	1·017	1·017
24	1·011	1·012	1·012	1·012	1·013
30	1·008	1·008	1·008	1·009	1·009
40	1·005	1·005	1·005	1·005	1·005
60	1·002	1·002	1·002	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	69·9185	74·4683	78·5672	83·5134	86·9938

$p = 8, r = 8$

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·491	1·538	1·585	1·646	1·692
2	1·270	1·288	1·305	1·326	1·342
3	1·185	1·195	1·204	1·215	1·224
4	1·138	1·144	1·150	1·158	1·163
5	1·108	1·113	1·117	1·123	1·126
6	1·088	1·091	1·095	1·099	1·102
7	1·073	1·076	1·078	1·082	1·084
8	1·061	1·064	1·066	1·069	1·071
9	1·053	1·055	1·057	1·059	1·060
10	1·046	1·048	1·049	1·051	1·052
12	1·036	1·038	1·039	1·040	1·041
14	1·028	1·030	1·031	1·031	1·032
16	1·023	1·025	1·026	1·026	1·027
18	1·020	1·021	1·022	1·022	1·023
20	1·017	1·017	1·018	1·018	1·018
24	1·012	1·013	1·013	1·014	1·014
30	1·009	1·009	1·009	1·009	1·010
40	1·005	1·005	1·006	1·006	1·006
60	1·003	1·003	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	78·8596	83·6753	88·0041	93·2169	96·8781

Table 7.2. (cont.)

$p = 3$, r = 10

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·650	1·716	1·781	1·862	1·921
2	1·333	1·359	1·383	1·413	1·435
3	1·218	1·232	1·245	1·262	1·274
4	1·157	1·167	1·175	1·187	1·195
5	1·120	1·127	1·133	1·141	1·147
6	1.095	1.101	1.106	1.112	1.116
7	1.078	1.082	1.086	1.091	1.095
8	1.065	1.068	1.072	1.075	1.078
9	1.066	1.058	1.061	1.064	1.066
10	1.047	1.050	1.051	1.055	1.057
12	1·036	1·038	1·040	1·042	1·043
14	1·029	1·030	1·031	1·033	1·034
16	1·023	1·024	1·025	1·027	1·028
18	1·019	1·020	1·021	1·022	1·023
20	1·016	1·017	1·018	1·019	1·019
24	1·012	1·012	1·013	1·014	1·014
30	1·008	1·009	1·009	1·009	1·010
40	1·005	1·005	1·005	1·006	1·006
60	1·002	1·002	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	40·2560	43·7730	46·9792	50·8922	53·6720

$p = 5$, r = 10

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·547	1·600	1·653	1·721	1·772
2	1·295	1·315	1·334	1·359	1·377
3	1·199	1·211	1·221	1·235	1·244
4	1·147	1·155	1·162	1·171	1·177
5	1·115	1·120	1·125	1·131	1·136
6	1.092	1.097	1.101	1.105	1.109
7	1.076	1.080	1.083	1.087	1.089
8	1.064	1.067	1.070	1.073	1.076
9	1.055	1.057	1.059	1.0	
10	1.048	1.050	1.051	1.0	
12	1·037	1·038	1·040	1·041	1·043
14	1·029	1·031	1·032	1·033	1·034
16	1·024	1·025	1·026	1·027	1·028
18	1·020	1·021	1·022	1·022	1·023
20	1·017	1·018	1·018	1·019	1·019
24	1·013	1·013	1·014	1·014	1·014
30	1·009	1·009	1·009	1·010	1·010
40	1·005	1·005	1·006	1·006	1·006
60	1·003	1·003	1·003	1·003	1·003
120	1·001	1·001	1·001	1·002	1·002
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	63·1671	67·5048	71·4202	76·1539	79·4900

$p = 7$, r = 10

$M \backslash \alpha$	0·100	0·050	0·025	0·010	0·005
1	1·509	1·557	1·604	1·666	1·713
2	1·285	1·303	1·320	1·342	1·359
3	1·197	1·208	1·217	1·229	1·238
4	1·148	1·155	1·162	1·169	1·175
5	1·117	1·122	1·127	1·132	1·136
6	1·096	1·099	1·103	1·107	1·110
7	1·080	1·083	1·086	1·089	1·091
8	1·068	1·070	1·073	1·075	1·077
9	1·058	1·060	1·062	1·065	1·066
10	1·051	1·053	1·054	1·056	1·058
12	1·040	1·042	1·042	1·044	1·045
14	1·032	1·034	1·034	1·036	1·036
16	1·026	1·028	1·028	1·029	1·029
18	1·022	1·023	1·023	1·024	1·024
20	1·019	1·019	1·020	1·020	1·021
24	1·014	1·014	1·015	1·015	1·016
30	1·010	1·010	1·010	1·011	1·011
40	1·006	1·006	1·006	1·007	1·007
60	1·003	1·003	1·003	1·003	1·003
120	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	85·5271	90·5312	95·0231	100·4250	104·2150

For p=4,q=10 enter the table
for p=10,r=4 with M=N-s-3
For p=6,r=10 enter the table
for p=10,r=6 with M=N-s-5

Source: M.Schatzoff (1966), Biometrika, pp.347-358

Table 7.3.

Chi-square adjustments to Wilks's criterion U. Factor C for lower percentiles of U (upper percentiles of χ^2_{pr})

	p=3, r=12					p=3, r=14				
α / M	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·718	1·791	1·860	1·949	2·013	1·780	1·857	1·931	2·026	2·095
2	1·382	1·410	1·437	1·470	1·495	1·427	1·458	1·486	1·523	1·549
3	1·256	1·272	1·287	1·306	1·319	1·292	1·309	1·326	1·346	1·361
4	1·188	1·199	1·209	1·221	1·230	1·217	1·229	1·240	1·254	1·264
5	1·146	1·154	1·161	1·170	1·176	1·171	1·179	1·188	1·198	1·205
6	1·117	1·123	1·129	1·136	1·141	1·138	1·145	1·152	1·159	1·165
7	1·097	1·101	1·106	1·111	1·115	1·115	1·121	1·126	1·132	1·136
8	1·081	1·085	1·089	1·093	1·097	1·097	1·102	1·106	1·111	1·115
9	1·069	1·073	1·076	1·080	1·082	1·084	1·088	1·091	1·095	1·099
10	1·060	1·063	1·066	1·069	1·071	1·073	1·076	1·079	1·082	1·085
12	1·046	1·048	1·050	1·053	1·054	1·057	1·059	1·061	1·064	1·066
14	1·037	1·039	1·040	1·042	1·043	1·046	1·048	1·049	1·052	1·053
16	1·030	1·032	1·033	1·034	1·035	1·037	1·039	1·041	1·042	1·044
18	1·025	1·026	1·027	1·029	1·029	1·031	1·033	1·034	1·035	1·036
20	1·021	1·022	1·023	1·024	1·025	1·027	1·028	1·029	1·030	1·031
24	1·016	1·017	1·017	1·018	1·019	1·020	1·021	1·022	1·023	1·023
30	1·011	1·011	1·012	1·012	1·013	1·014	1·015	1·015	1·016	1·016
40	1·007	1·007	1·007	1·008	1·008	1·009	1·009	1·009	1·010	1·010
60	1·003	1·003	1·004	1·004	1·004	1·004	1·004	1·005	1·005	1·005
120	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	47·2122	50·9985	54·4373	58·6192	61·5812	54·0902	58·1240	61·7768	66·2062	69·3360

	r=16					r=18				
α / M	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·835	1·916	1·995	2·095	2·169	1·886	1·971	2·053	2·158	2·235
2	1·469	1·501	1·532	1·571	1·599	1·508	1·542	1·575	1·616	1·646
3	1·325	1·344	1·362	1·384	1·400	1·357	1·377	1·396	1·420	1·437
4	1·245	1·258	1·271	1·285	1·296	1·272	1·286	1·299	1·315	1·327
5	1·195	1·204	1·213	1·224	1·232	1·218	1·228	1·238	1·249	1·258
6	1·159	1·167	1·174	1·182	1·188	1·179	1·188	1·195	1·204	1·211
7	1·133	1·139	1·145	1·152	1·157	1·151	1·158	1·164	1·171	1·177
8	1·114	1·119	1·123	1·129	1·133	1·129	1·135	1·140	1·146	1·151
9	1·098	1·102	1·106	1·111	1·115	1·112	1·117	1·121	1·127	1·130
10	1·085	1·089	1·092	1·097	1·099	1·099	1·103	1·107	1·111	1·114
12	1·067	1·070	1·073	1·076	1·078	1·078	1·081	1·084	1·087	1·090
14	1·054	1·057	1·059	1·061	1·063	1·064	1·066	1·068	1·071	1·073
16	1·045	1·047	1·049	1·051	1·052	1·053	1·055	1·057	1·059	1·061
18	1·038	1·039	1·041	1·043	1·044	1·045	1·046	1·048	1·050	1·051
20	1·032	1·034	1·035	1·036	1·037	1·038	1·040	1·041	1·043	1·044
24	1·025	1·026	1·026	1·027	1·028	1·029	1·030	1·031	1·032	1·033
30	1·017	1·018	1·018	1·019	1·020	1·021	1·021	1·022	1·023	1·023
40	1·011	1·011	1·011	1·012	1·012	1·013	1·013	1·014	1·014	1·015
60	1·005	1·006	1·006	1·006	1·006	1·006	1·007	1·007	1·007	1·007
120	1·002	1·002	1·002	1·002	1·002	1·002	1·002	1·002	1·002	1·002
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	60·9066	65·1708	69·0226	73·6826	76·9688	67·6728	72·1532	76·1920	81·0688	84·5019

Table 7.3. (cont.)

	p=3,r=20					p=3,r=22				
$M \diagdown \alpha$	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·932	2·021	2·106	2·216	2·297	1·975	2·067	2·156	2·269	2·353
2	1·544	1·580	1·614	1·657	1·689	1·578	1·616	1·651	1·696	1·729
3	1·387	1·408	1·428	1·453	1·472	1·415	1·438	1·459	1·485	1·504
4	1·298	1·313	1·327	1·344	1·356	1·322	1·338	1·353	1·371	1·384
5	1·240	1·251	1·261	1·274	1·283	1·261	1·273	1·284	1·297	1·307
6	1·199	1·208	1·216	1·226	1·233	1·218	1·227	1·236	1·246	1·254
7	1·168	1·176	1·182	1·190	1·196	1·185	1·193	1·200	1·209	1·215
8	1·145	1·151	1·157	1·163	1·168	1·160	1·167	1·173	1·180	1·185
9	1·127	1·132	1·136	1·142	1·146	1·142	1·147	1·151	1·157	1·161
10	1·112	1·116	1·120	1·125	1·128	1·124	1·129	1·133	1·139	1·141
12	1·089	1·092	1·095	1·099	1·102	1·099	1·103	1·106	1·110	1·113
14	1·073	1·075	1·078	1·081	1·083	1·082	1·085	1·087	1·091	1·093
16	1·061	1·063	1·065	1·067	1·069	1·069	1·071	1·073	1·076	1·078
18	1·052	1·053	1·055	1·057	1·059	1·059	1·061	1·063	1·065	1·066
20	1·044	1·046	1·048	1·049	1·050	1·051	1·052	1·054	1·056	1·057
24	1·034	1·035	1·036	1·038	1·039	1·039	1·040	1·041	1·043	1·044
30	1·024	1·025	1·026	1·027	1·027	1·028	1·029	1·030	1·031	1·031
40	1·015	1·016	1·016	1·017	1·017	1·018	1·018	1·019	1·020	1·020
60	1·008	1·008	1·008	1·009	1·009	1·009	1·009	1·010	1·010	1·010
120	1·002	1·002	1·002	1·002	1·003	1·003	1·003	1·003	1·003	1·003
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	74·3970	79·0819	83·2976	88·3794	91·9517	81·0855	85·9649	90·3489	95·6257	99·3304

Table 7.3 (cont.)

$M \backslash \alpha$	p=4, r=12					p=4, r=14				
	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·638	1·700	1·760	1·838	1·895	1·686	1·751	1·814	1·896	1·956
2	1·350	1·373	1·396	1·424	1·446	1·388	1·413	1·436	1·467	1·489
3	1·238	1·252	1·264	1·280	1·292	1·269	1·284	1·297	1·314	1·327
4	1·177	1·186	1·195	1·205	1·213	1·203	1·213	1·222	1·234	1·242
5	1·139	1·145	1·152	1·159	1·165	1·161	1·168	1·175	1·183	1·189
6	1·112	1·118	1·122	1·128	1·132	1·131	1·137	1·142	1·149	1·154
7	1·093	1·097	1·101	1·106	1·109	1·110	1·115	1·119	1·124	1·128
8	1·079	1·082	1·085	1·089	1·092	1·094	1·097	1·101	1·105	1·109
9	1·068	1·070	1·073	1·076	1·079	1·081	1·084	1·087	1·091	1·093
10	1·059	1·061	1·063	1·066	1·068	1·071	1·073	1·076	1·079	1·081
12	1·046	1·047	1·049	1·051	1·053	1·055	1·058	1·059	1·062	1·064
14	1·037	1·038	1·039	1·041	1·042	1·045	1·046	1·048	1·050	1·051
16	1·030	1·031	1·032	1·033	1·034	1·037	1·038	1·039	1·041	1·042
18	1·025	1·026	1·027	1·028	1·029	1·031	1·032	1·033	1·034	1·035
20	1·021	1·022	1·023	1·024	1·024	1·026	1·027	1·028	1·029	1·030
24	1·016	1·017	1·017	1·018	1·018	1·020	1·021	1·021	1·022	1·023
30	1·011	1·011	1·012	1·012	1·013	1·014	1·014	1·015	1·015	1·016
40	1·007	1·007	1·007	1·008	1·008	1·009	1·009	1·009	1·009	1·010
60	1·003	1·003	1·004	1·004	1·004	1·004	1·004	1·005	1·005	1·005
120	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	60·9066	65·1708	69·0226	73·6826	76·9688	69·9185	74·4683	78·5671	83·5134	86·9937

$M \backslash \alpha$	p=4, r=16					p=4, r=18				
	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·731	1·799	1·864	1·949	2·012	1·773	1·843	1·911	1·999	2·065
2	1·423	1·450	1·475	1·507	1·531	1·457	1·485	1·511	1·545	1·570
3	1·299	1·314	1·329	1·347	1·360	1·327	1·343	1·359	1·378	1·392
4	1·228	1·239	1·249	1·261	1·270	1·252	1·264	1·274	1·287	1·297
5	1·182	1·190	1·198	1·207	1·213	1·203	1·212	1·220	1·230	1·237
6	1·150	1·157	1·163	1·169	1·174	1·169	1·176	1·182	1·189	1·195
7	1·127	1·132	1·136	1·142	1·146	1·143	1·149	1·154	1·160	1·164
8	1·108	1·113	1·117	1·121	1·125	1·123	1·128	1·132	1·137	1·141
9	1·094	1·098	1·101	1·105	1·108	1·107	1·111	1·115	1·119	1·122
10	1·083	1·086	1·089	1·092	1·094	1·095	1·098	1·101	1·105	1·108
12	1·065	1·068	1·070	1·073	1·074	1·075	1·078	1·080	1·083	1·085
14	1·053	1·055	1·056	1·058	1·060	1·061	1·063	1·065	1·068	1·069
16	1·044	1·045	1·047	1·049	1·050	1·051	1·053	1·054	1·056	1·058
18	1·037	1·038	1·040	1·041	1·042	1·044	1·045	1·046	1·048	1·049
20	1·032	1·033	1·034	1·035	1·036	1·037	1·039	1·040	1·041	1·042
24	1·024	1·025	1·026	1·027	1·027	1·029	1·030	1·030	1·031	1·032
30	1·017	1·018	1·018	1·019	1·019	1·020	1·021	1·022	1·022	1·023
40	1·011	1·011	1·011	1·012	1·012	1·013	1·013	1·014	1·014	1·014
60	1·005	1·005	1·006	1·006	1·006	1·006	1·007	1·007	1·007	1·007
120	1·001	1·002	1·002	1·002	1·002	1·002	1·002	1·002	1·002	1·002
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	78·8597	83·6753	88·0040	93·2168	96·8781	87·7431	92·8083	97·3531	102·816	106·648

Table 7.3 (cont.)

	p=5, r=12		p=5, r=14		p=5, r=16	
$M \diagdown \alpha$	0·050	0·010	0·050	0·010	0·050	0·010
1	1·643	1·768	1·683	1·813	1·722	1·855
2	1·350	1·396	1·383	1·431	1·415	1·465
3	1·240	1·265	1·267	1·294	1·294	1·323
4	1·179	1·196	1·203	1·221	1·226	1·245
5	1·141	1·153	1·161	1·174	1·181	1·196
6	1·114	1·124	1·132	1·143	1·150	1·161
7	1·095	1·103	1·111	1·119	1·127	1·136
8	1·081	1·087	1·095	1·102	1·109	1·116
9	1·070	1·075	1·082	1·088	1·095	1·101
10	1·061	1·065	1·072	1·077	1·083	1·089
12	1·047	1·051	1·057	1·060	1·066	1·070
14	1·038	1·040	1·045	1·048	1·053	1·057
16	1·031	1·033	1·038	1·040	1·045	1·047
18	1·026	1·028	1·032	1·034	1·038	1·040
20	1·022	1·024	1·027	1·029	1·033	1·034
24	1·017	1·018	1·021	1·022	1·025	1·026
30	1·012	1·012	1·014	1·015	1·018	1·019
40	1·007	1·008	1·009	1·010	1·011	1·012
60	1·004	1·004	1·004	1·005	1·006	1·006
120	1·001	1·001	1·001	1·001	1·002	1·002
∞	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	79·0819	88·3794	90·5312	100·4250	101·8790	112·3290

	p=6, r=11		p=6, r=12		p=6, r=13	
$M \diagdown \alpha$	0·050	0·010	0·050	0·010	0·050	0·010
1	1·589	1·704	1·605	1·722	1·621	1·739
2	1·321	1·363	1·335	1·378	1·349	1·393
3	1·220	1·243	1·232	1·255	1·244	1·268
4	1·164	1·180	1·175	1·191	1·185	1·201
5	1·129	1·140	1·138	1·150	1·148	1·159
6	1·105	1·114	1·113	1·122	1·121	1·130
7	1·088	1·094	1·095	1·102	1·102	1·109
8	1·074	1·080	1·081	1·086	1·088	1·093
9	1·064	1·069	1·070	1·075	1·075	1·080
10	1·055	1·060	1·061	1·065	1·066	1·070
12	1·044	1·047	1·048	1·051	1·052	1·055
14	1·035	1·037	1·038	1·040	1·042	1·044
16	1·029	1·030	1·032	1·033	1·035	1·037
18	1·024	1·026	1·027	1·028	1·029	1·031
20	1·020	1·022	1·023	1·024	1·025	1·026
24	1·015	1·016	1·017	1·018	1·019	1·020
30	1·011	1·011	1·012	1·013	1·013	1·014
40	1·007	1·007	1·007	1·008	1·008	1·009
60	1·003	1·003	1·004	1·004	1·004	1·004
120	1·001	1·001	1·001	1·001	1·001	1·001
∞	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	85·9649	95·6257	92·8083	102·8160	99·6169	109·9580

Table 7.3 (cont.)

	p=4, r=20					p=4, r=22				
α / M	0·100	0·050	0·025	0·010	0·005	0·100	0·050	0·025	0·010	0·005
1	1·812	1·884	1·954	2·045	2·113	1·848	1·922	1·994	2·088	2·158
2	1·488	1·518	1·545	1·580	1·606	1·518	1·549	1·577	1·614	1·641
3	1·353	1·371	1·387	1·408	1·422	1·379	1·397	1·414	1·436	1·451
4	1·275	1·288	1·299	1·313	1·323	1·298	1·310	1·322	1·337	1·347
5	1·224	1·233	1·241	1·252	1·259	1·243	1·253	1·262	1·273	1·281
6	1·187	1·194	1·201	1·208	1·215	1·204	1·212	1·219	1·228	1·234
7	1·159	1·165	1·170	1·177	1·182	1·175	1·181	1·187	1·194	1·199
8	1·138	1·143	1·147	1·153	1·157	1·152	1·157	1·162	1·168	1·172
9	1·121	1·125	1·129	1·133	1·137	1·134	1·138	1·142	1·147	1·151
10	1·107	1·110	1·114	1·118	1·121	1·119	1·123	1·126	1·130	1·134
12	1·086	1·088	1·091	1·094	1·096	1·095	1·098	1·101	1·104	1·107
14	1·070	1·072	1·074	1·077	1·078	1·079	1·081	1·083	1·086	1·088
16	1·059	1·061	1·062	1·064	1·066	1·066	1·068	1·070	1·072	1·074
18	1·050	1·052	1·053	1·055	1·056	1·057	1·058	1·060	1·062	1·063
20	1·043	1·045	1·046	1·047	1·048	1·049	1·051	1·052	1·053	1·055
24	1·033	1·034	1·035	1·036	1·037	1·038	1·039	1·040	1·041	1·042
30	1·024	1·024	1·025	1·026	1·026	1·027	1·028	1·029	1·030	1·030
40	1·015	1·016	1·016	1·016	1·017	1·017	1·018	1·018	1·019	1·019
60	1·008	1·008	1·008	1·008	1·008	1·009	1·009	1·009	1·010	1·010
120	1·002	1·002	1·002	1·002	1·002	1·003	1·003	1·003	1·003	1·003
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
χ^2_{pr}	96·5782	101·879	106·629	112·329	116·321	105·372	110·898	115·841	121·767	125·913

Source: K.C.S. Pillai and A.K. Gupta (1969). Biometrika, pp.109-118.

EXERCISE 7

1. The following data relating to the first lactation yield (x_1, x_2) in 10 lb units of dams (x_1) and of daughters (x_2) of 5 sires are taken from the records of Kankrej herd maintained at the Institute of Agriculture, Anand, India (Source: National index of field experiments M.P.6, Indian Agricultural Research Institute, New Delhi, India).

Sire 1	Sire 2	Sire 3	Sire 4	Sire 5
216,258	153,112	274,313	340,340	152,180
216,178	276,269	375,374	251,347	375,236
269,243	260,247	312,320	169,193	198,120
236,178	299,198	229,532	208,235	242,140
156,111	298,357	171,203	182,252	269,227
280,233	266,327	197,269	221,280	221,140
190, 60	281,200	226,153	193,194	532,277
340,262	186,190	260,228	312,169	187,179
216,191	188,261	166,184	299,242	
277,138	168,235	252,383	188,253	
121,197	251,165			
261,234				
190,123				

Analyse the data to test the hypothesis that the population means pertaining to different sires are equal.

2. The following bivariate data (x_1, x_2), x_1 = total green weight in lbs. of jute plants per plot, x_2 = total dry fibre weight of jute per plot were collected from a fertilizer trial, with 12 different combinations of (a) 3 levels: $P_0 = 0$ lb per acre, $P_1 = 40$ lb per acre, $P_2 = 80$ lb per acre of P_2O_5 as superphosphate, (b) 2 levels: $N_o = 0$ lb per acre, $N_1 = 20$ lb per acre of nitrogen as ammonium sulphate, (c) 2 levels: $K_0 = 0$ lb per acre $K_1 = 25$ lb per acre of K_2O as murate of potash, on a particular variety of jute crop. The experiment was conducted in Sagar-Bishnupur, West Bengal, India and the layout was a randomized block design four blocks of 18'×19' (net) each.

	Blocks			
Treatments	1	2	3	4
$P_0N_0K_0$	88.3,5.66	104.9,7.42	106.3,7.17	101.8,6.70
$P_1N_0K_0$	93.4,6.31	112.4,8.34	140.0,10.0	111.6,7.57
$P_2N_0K_0$	94.1,6.72	110.5,8.24	109.8,7.40	112.4,7.36
$P_0N_0K_1$	96.6,6.66	106.1,7.60	98.2,6.50	92.6,6.24
$P_0N_1K_0$	92.2,6.39	104.6,7.18	104.8,7.00	115.8,7.95
$P_1N_0K_1$	103.1,7.32	105.9,7.60	135.3,8.90	115.0,7.85
$P_1N_1K_0$	101.5,7.17	111.6,8.00	106.2,7.40	118.3,7.90
$P_2N_1K_0$	67.1,5.04	105.0,7.33	120.0,8.24	104.9,6.68
$P_0N_1K_1$	90.1,6.10	104.9,7.44	106.8,7.06	94.0,6.35
$P_1N_1K_1$	110.2,8.06	107.2,7.51	106.1,7.13	104.9,7.10
$P_2N_1K_1$	102.9,7.25	105.9,7.72	143.6,8.91	106.1,7.06
$P_2N_0K_1$	100.1,6.82	121.3,8.10	110.1,7.40	121.0,8.50

Analyze the data to examine if the different fertilizers have different effects on this particular variety of Jute.

REFERENCES

Anderson, T.W. (1958). An introduction to multivariate analysis, Wiley, New York.

Anderson, T.W., Das Gupta, S. and Mudolkar, G.S. (1964). Monotonicity of power functions of some tests of multivariate general linear hypothesis, Ann.Math.Statist., 35, 200-205.

Bartlett, M.S. (1934). The vector representation of a sample. Proc. R. Soc. Edinburgh, 53, 260-283.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood ratio criteria. Biometrika 36, pp.317-346.

Consul, P.C. (1966). On the exact distributions of likelihood ratio criteria for testing linear hypothesis about regression coefficients. Ann.Math.Statist. 37, 1319-1330.

Foster, R.G. (1957). Upper percentage point of the generalized beta distribution II, Biometrika, 44, 441-453.

Foster, R.G. (1958). Upper percentage point of the generalized beta distribution III, Biometrika 45, 492-503.

Foster, R.G. and Rees, D.D. (1957). Upper percentage point of the generalized beta distribution I, Biometrika, 44, 237-247.

Gabriel, K.R. (1964). Comparison of some methods of simultaneous inference in MANOVA in multivariate analysis (P.R.Krishnaiah ed.) Vol.II, 67-86. Academic Press, New York.

Ghosh, M.N. (1964). On the admissibility of some tests of MANOVA, Ann. Math. Statist. 35, 789-794.

Giri, N. (1976). Invariance and minimax statistical tests, Hindustan Publishing Corp., New Delhi.

Giri, N. (1977). Multivariate statistical inference, Academic Press, New York.

Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax property of multivariate tests, Ann. Math. Statist. 35, pp.21-35.

Giri, N., Kiefer, J. and Stein, C. (1963). Minimax character of T^2-test in the simplest case, Ann. Math. Statist. 34, 1524-1535.

Hannan, E.J. (1970). Multiple time series, Wiley, New York.

Heck, D.L. (1960). Charts of some upper percentage points of the distribution of the largest characteristic root. Ann. Math. Statist., 31 625-642.

Hotelling, H. (1951). A generalized T-test and measure of multivariate dispersion. Proc. Berkeley Symp. Prob. Statist. 2nd, 23-41.

Hsu, P.L. (1940). On a generalized analysis of variance I, Biometrika, 31, 221-237.

James, A.T. (1964). Distribution of matrix variates and latent roots derived from normal samples, Ann. Math. Statist., 35, 475-501.

John, S. (1971). Some optimal multivariate tests, Biometrika, 58, 123-127.

Khatri, C.G. (1964a). Distribution of the generalized multiple correlation matrix in the dual case. Ann. Math. Statist. 35, 1801-1805.

Khatri, C.G. (1964b). Distribution of the largest and the smallest characteristic root under null hypothesis concerning complex multivariate normal populations. Ann.Math.Statist. 35,1806-1810.

Khatri, C.G. (1970). On the moments of traces of two matrices in three situations for complex multivariate normal populations, Sankhya, 32, 65-80.

Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2-, R^2 and other fully invariant tests for classical normal problems, Ann. Math. Statist. 36, 747-760.

Lawley, D.N. (1938). A generalization of Fisher's Z-test, Biometrika, 30, 180-187.

Lee, Y.S. (1972). Some results on the distribution of Wilks' likelihood criterion, Biometrika, 59, 649-664.

Lehmann, E. (1959). Testing statistical hypotheses, Wiley, New York.

Linnik, Ju, Vo., Pliss, V.A. and Salaevskie, O.V. (1966). Dokl. Adad. Nauk. SSSR 168.

Mathai, A.M. (1971a). On the distribution of likelihood ratio criterion for testing linear hypotheses on regression coefficients, Ann. Inst. Statist. Math. 23, 181-197.

Mathai, A.M. (1971b). An expansion of Meijer's G-function and the distribution of products of independent beta variates, South African Statist. J., 5, 71-90.

Mathai, A.M. and Rathie, P.N. (1971). The exact distribution of Wilks' criterion, Ann. Math. Statist. 41, 53-65.

Marrow, D.J. (1948). On the distribution of the sums of characteristic roots of a determinantal equation (Abstract), Bull. Am. Math. Soc., 54, p. 75.

Nair, U.S. (1938). The application of the moment function in the study of distribution laws in statistics, Biometrika, 30, 274-294.

Nandi, H.K. (1963). Admissibility of a class tests, Calcutta Statist. Assoc. Bull., 15, 13-18.

Narain, R.D. (1950). On the completely unbiased character of tests of independence in multivariate normal system, Ann. Math. Statist., 21, 293-298.

Olkin, I. (1952). Note on the Jacobian of some matrix transformations useful in multivariate analysis, Biometrika,40,43-46.

Pillai, K.C.S. (1957). Concise tables of statisticians, the statistical center, University of the Philippines, Manila.

Pillai, K.C.S. (1960). Statistical tables for tests of multivariate hypotheses, The statistical center, University of the Philippines, Manila.

Pillai, K.C.S. (1976). Distribution of characteristic roots, (Invited paper), Canadian Journal of Statistics.

Pillai, K.C.S. and Gupta, A.K. (1969). On the exact distribution of Wilks' criterion, Biometrika, 56, 109-118.

Pillai, K.C.S. and Jayachandran, K. (1970). On the exact distribution of Pillai's $V^{(s)}$ criterion, J. Amer. Statist. Assoc. 65, 447-454.

Pillai, K.C.S. and Jouris, G.M. (1972). An approximation to the distribution of the largest root of a matrix in the complex Gaussian case, Ann. Inst. Statist. Math. 24, pp. 517-525.

Pillai, K.C.S. and Young, D.L. (1970). An approximation to the distribution of the largest root of a matrix in the complex Gaussian case, Ann. Inst. Statist. Math., 22, 89-96.

Rao, C.R. (1951). An asymptotic expansion of the distribution of Wilks' Λ-criterion, Bull. Inst. Internat. Statist. 33, II, 177-180.

Rao, C.R. (1952). Advanced statistical methods in biometric research, Wiley, New York.

Rao, C.R. (1965). Linear statistical inference and its applications, Wiley, New York.

Roy, S.N. (1957). Some aspects of multivariate analysis, Wiley, New York.

Roy, S.N. and Bose, R.C. (1953). Simultaneous confidence interval estimation, Ann. Math. Statist., 24, 513-536.

Roy, S.N. and Mikhai, W.F. (1961). On the monotonic character of the power functions of two multivariate tests, Ann. Math. Statist. 32, 1145-1151.

Salaevskii, O.V. (1968). Minimax character of Hotelling's T^2-test, Sov.Math.Dokl.9, 733-735.

Schatzoff, M. (1966). Exact distribution of Wilks' likelihood ratio crtierion, Biometrika, 53, 347-358.

Schwartz, R. (1964a). Properties of tests in MANOVA (Abstract), Ann. Math. Statist., 35, 939.

Schwartz, R. (1964b). Admissible invariant test in MANOVA (Abstract), Ann. Math. Statist., 35, 1398.

Sinha, B. and Giri, N. (1977). On the optimality and non-optimality of some multivariate test procedures, (in press), Sankhya.

Smith, H., Gnanadeshikhan, R. and Huges, J.B. (1962). Multivariate analysis of variance, Biometrika, 18, 22-41.

Stein, C. (1956). The admissibility of Hotelling's T^2-test. Ann. Math. Statist. 27, 616-623.

Wilks, S.S. (1932). Certain generalization of analysis of variance, Biometrika, 24, 471-474.

Wilks, S.S. (1934). Moment generating operator for determinant product moments in samples from normal system, Ann. Math. Statist. 5, 312-340.

Wilks, S.S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses,Ann.Math.Statist. 9,101-112.

CHAPTER 8. NONPARAMETRIC ANALYSIS OF VARIANCE

Nonparametric procedures are essentially concerned with the development of statistical methods without making any explicit assumption regarding the functional form of the probability distributions of the sample observations. Uptil now we were concerned with cases where the functional form is known to be univariate or multivariate normal. In this chapter we shall deal with cases where wo do not make any explicit assumption about the functional form of the univariate sample observations. The multivariate analogue of these results are more complicated and we refer to Puri and Sen (1971) for multivariate procedures analogues to the univariate ones discussed in this chapter.

The relative performance of various nonparametric methods are usually studied for large samples and the study of their small sample properties often involves quite cumbersome computational scheme. The development of the asymptotic theory of nonparametric statistical inference is somewhat spotty and piecemeal as compared to its parametric counterparts. We refer to Puri and Sen (1971) and Lehmann (1975) for further readings and complete references on this topic. In our development of univariate nonparametric analysis of variance we mainly state various results without proof and we refer to above references for their proof. Before we start discussing the one-way and higher-way classifications we deal briefly some relevant results on the estimation in one-sample and two-sample problems and testing of hypotheses of two-sample location parameters.

8.1. ONE-SAMPLE LOCATION PARAMETER (ESTIMATION)

Suppose that $X_1,\dots,X_n$ are distributed according to a common symmetric distribution $F(x-\theta)$ whose centre of symmetry θ is to be estimated. Let us assume that F is absolutely continuous. When F is normal with finite variance it is well known that

$$\overline{\theta} = \frac{1}{n} \sum_{1}^{n} X_i \tag{8.1}$$

is the minimum variance unbiased estimator of θ. This estimator is very sensitive to gross errors and tends to be inefficient with heavy tails. An example of this fact is provided by the Cauchy distribution for which the distribution of $\overline{\theta}$ for any n is the same as X_1 so that the average is not better than the single observation. In view of this fact that the distribution of $X_1-\theta,\ldots,X_n-\theta$ is symmetric about the origin, the best estimator of θ is that value of θ for which the values $X_1-\theta,\ldots,X_n-\theta$ give the best balance relative to the origin. Consider a test statistic $h(X_1,\ldots,X_n)$ for testing the hypothesis that the common distribution of $X_1-\theta,\ldots,X_n-\theta$ is symmetric about the origin. Obviously the variables $X_1-\theta,\ldots,X_n-\theta$ will be best balanced about the origin when the test statistic h evaluated for $X_1 - \theta,\ldots, X_n - \theta$ takes on the value giving the strongest support to the hypothesis mentioned above. This will typically be the case when h takes on its central value. One may thus expect the resulting estimator to share some of the properties of the test from which it is derived. A few commonly used test statistics are as follows.

(i) Sign-test statistics: Evaluated for the variables $X_1-\theta,\ldots,X_n-\theta$ it is the number among the $X_i-\theta$'s which are positive. Obviously, it gives the greatest support to the hypothesis, that the variables $X_1-\theta,\ldots,X_n-\theta$ are symmetric about 0, if half of the variables $X_i-\theta$ are positive and half are negative and it happens if θ is the median of the X_i. So the sign test leads to the estimate

$$\tilde{\theta} = \text{med}(X_i) \tag{8.2}$$

where $\text{med}(X_i)$ denotes the median of the X_i.

(ii) Wilcoxon signed rank test statistic: Evaluated for the variables it is the number of pairs (i,j) with $i \leq j$ for which

$X_i - \theta + X_j - \theta$ is positive and it takes on its central value if half of the sums $X_i - \theta + X_j - \theta$, for all $i \leq j$ are positive and half negative. Thus the corresponding estimator is the median of the $n(n+1)/2$ values $(X_i + X_j)/2$ for $1 \leq i \leq j \leq n$, that is,

$$\hat{\theta} = \underset{i \leq j}{\text{med}} \ \tfrac{1}{2}(X_i + X_j) \tag{8.3}$$

(see Hodges and Lehmann (1963), Sen (1963) for details).

Example 8.1. Consider the sample of size 6 with values 7.4, 5.4, 6.2, 7.6, 5.8, 5.4, from a distribution. It is desired to estimate the centre θ of the distribution which is assumed to be symmetric about θ. From the sign-test statistic $\tilde{\theta} = \frac{1}{2}(5.8 + 6.2) = 6.0$. To get $\hat{\theta}$ we first observe that there are 21 sums $\frac{1}{2}(X_i + X_j)$ with $i \leq j$ and the median of these sums will be the 11th smallest. On ordering the observations we rearrange the sample as 5.1, 5.4, 5.8, 6.2, 7.4, 7.6. So the 11th smallest of the sums is given by $\frac{1}{2}(5.4 + 7.6) = 6.5$. Since

$$\hat{\theta} - \theta = \text{med}(\tfrac{1}{2}(X_i + X_j) - \theta)$$

$$= \text{med}(\tfrac{1}{2}(X_i - \theta) + \tfrac{1}{2}(X_j - \theta))$$

and the distribution of $X_i - \theta$ and $X_j - \theta$ is independent of θ, $\hat{\theta} - \theta$ is independent of θ. Similar results hold for $\bar{\theta} - \theta$ and $\tilde{\theta} - \theta$. Furthermore, from problem 1 if the distribution of $X_1, \ldots, X_n$ is symmetric about θ the same is true about the distribution of $\bar{\theta}$, $\tilde{\theta}$ and $\hat{\theta}$. This result shows that each of the three estimators are unbiased for θ and each is as likely to overestimate θ as to underestimate it. It is known that (see for example Lehmann(1975)) $\hat{\theta}$ and $\bar{\theta}$ are considerably more efficient than $\tilde{\theta}$ when F is normal with then $\bar{\theta}$ being slightly more efficient than $\hat{\theta}$. But for distributions with tails heavier than the normal $\hat{\theta}$ may be considerably more efficient than $\bar{\theta}$ and in general it appears to be a reasonable one. On the other hand when the observations can easily be ordered then $\tilde{\theta}$ can be computed easily and is usually used when F is not symmetric. Furthermore if F has finite expectation then $\bar{\theta}$ is a consistent estimator of θ and if F has a unique median θ then $\tilde{\theta}$ is a consistent estimator of θ. If θ be such that

$$P\{\tfrac{1}{2}(X+Y) < \theta\} = P\{\tfrac{1}{2}(X+Y) > \theta\} = \tfrac{1}{2}$$

where X, Y are independently distributed as F and if θ is the only point with this property then $\hat{\theta}$ is consistent for θ.

8.2. TWO-SAMPLE LOCATION PARAMETERS (ESTIMATION AND TESTING)

Suppose that $X_1,\ldots,X_m; Y_1,\ldots,Y_n$, $N=m+n$, are independent samples from the absolutely continuous distribution functions F(x) and $G(x) = F(x-\theta)$. This is called an additive model (also shift model) which assumes that the treatment adds a constant amount θ to the control response of the subject independent of the value of that response. Since the variables $X_1,\ldots,X_m; Y_1-\theta,\ldots,Y_n-\theta$ have the same distribution we may then estimate θ by $\hat{\theta}$ – the amount by which the Y values have to be shifted to give the Y values the best possible agreement with the X-values. To achieve this let us consider a test statistic

$$h(X_1,\ldots,X_m; Y_1,\ldots,Y_n) = h(X,Y) \text{ (say)}$$

for testing $H_o: \theta=0$ against alternatives $\theta \neq 0$. Suppose that under H_o the test statistic h(X, Y) is symmetrically distributed about some value μ and let the two sided test rejects H_o whenever

$$h(X,Y) \geq \mu + c \quad \text{or} \quad h(X,Y) \leq \mu - c.$$

Then the value of h(X, Y) which is most supportive of H_o is the value closest to the centre μ of the distribution. The two sets $X_1,\ldots,X_m$ and $Y_1-\hat{\theta},\ldots,Y_n-\hat{\theta}$ will therefore be considered as being in the best agreement if they assign to the statistic $h(X_1,\ldots,X_m; Y_1-\hat{\theta},\ldots,Y_n-\hat{\theta})$ this closest value. A commonly used test statistic h(X, Y) is the Wilcoxon statistic $W_{X,Y}$. By definition $W_{X,Y-\theta}$ is the number of pairs (i, j) for which $Y_j-\theta > X_i$ or $Y_j - X_i > \theta$. So $W_{X,Y-\theta}$ will take on its central value if half of the differences Y_j-X_i are greater than θ and half are less than θ. Thus the value $\hat{\theta}$ of θ for which this is achieved is the median of the mn differences Y_j-X_i (see Hodges and Lehmann (1963) and Sen (1963) for details). Let $d_{(1)},\ldots,d_{(mn)}$ denote the ordered set of mn differences Y_j-X_i.

Case a. Let mn = 2k where k is an integer. Since the centre of symmetry of $W_{X,Y}$ is mn/2 = k, the value we are looking for a value of θ such that $W_{X,Y-\theta}=k$, that is, there are exactly k of the differences Y_j-X_i that exceeds θ. Obviously this is true if and only if $d_{(k)} < \theta < d_{(k+1)}$. So $W_{X,Y-\theta}$ takes the central value not for a single value of θ but for an interval of such values. Obviously then the natural estimator of θ is the midrank of the interval $(d_{(k)}, d_{(k+1)})$ and is given by

$$\hat{\theta} = \tfrac{1}{2}(d_{(k)}+d_{(k+1)}) = \mathrm{med}(Y_j-X_i) \tag{8.4}$$

Case b. mn = 2k+1. The centre of symmetry of the distribution of $W_{X,Y-\theta}$ is $k+\frac{1}{2}$. Clearly $k+\frac{1}{2}$ is not a value of the integer valued statistic $W_{X,Y-\theta}$. However it is easy to conclude that $W_{X,Y-\theta} > k+\frac{1}{2}$ if and only if $\theta < d_{(k+1)}$ and $W_{X,Y-\theta} < k+\frac{1}{2}$ if and only if $\theta > d_{(k+1)}$. Thus $W_{X,Y-\theta}$ jumps down from k+1 to k as θ moves through the point $d_{(k+1)}$ and if (X_i, Y_j) is the pair with $Y_j-X_i = d_{(k+1)}$ a tie occurs between X_i and $Y_j-\theta$. Taking the contribution of a tie to $W_{X,Y-\theta}$ to be $\frac{1}{2}$ it follows that the associated value of $W_{X,Y-\theta}=k+\frac{1}{2}$ and we take $\hat{\theta} = d_{(k+1)} = \mathrm{med}(Y_j-X_i)$.

It can be verified that in the shift model under the assumption that F is continuous the distribution $\hat{\theta}-\theta=\mathrm{med}(Y_j-X_i)-\theta$ is independent of θ. Also $\hat{\theta}$ is symmetrically distributed about θ if either the distribution F is symmetric about θ or if m=n and this implies that $E(\hat{\theta})=\theta$. Furthermore $\hat{\theta}$ is consistent for estimating θ when both $m,n \to \infty$.

We shall now consider an alternative expression for $W_{X,Y}$ in terms of a statistic R which is commonly known as the Mann-Whitney statistic.

Let us rank the m+n observations $X_1,\dots,X_m$, $Y_1,\dots,Y_n$ in the ascending order of magnitudes and let $R_1,\dots,R_n$ be the ranks of $Y_1,\dots,Y_n$. Denote $R=\sum_1^n R_i$ then (8.5)

$$W_{X,Y} = R - \frac{n(n+1)}{2}$$

To prove (8.5) let us assume without any loss of generality that $Y_1 < Y_2 < \dots < Y_n$. Since the rank of Y_1 is R_1 there are R_1-1 observations among the X_i each of which is smaller than Y_1. Similarly there are R_2-2 observations among the X_i each of which is smaller than Y_2 and so on. Hence

$$\begin{aligned} W_{X,Y} &= R_1-1+R_2-1+ \dots + R_n-n \\ &= R - \frac{n(n+1)}{2} \; . \end{aligned}$$

Let us now consider the problem of testing the hypothesis $H_o:\theta=0$. In deciding an appropriate test it is not enough to specify H_o alone. We should also specify the alternatives against which it is tested. The model

$$G(x) = F(x-\theta) \tag{8.6}$$

for all x expresses the fact that the distribution G is obtained by shifting F by an amount $\theta > 0$. In other words the model (8.6) supposes that the treatment adds a constant positive amount θ to the control response. In the shift model it is then natural to consider the alternative hypothesis H_1: $\theta > 0$. The Wilcoxon test of H_o against H_1 rejects H_o whenever

$$W_{X,Y} \geq C \tag{8.7}$$

where the constant C is determined so that the test has size α. It is known that the power function of the test against the alternatives H_1: $\theta > 0$ is a nondecreasing function of θ and it is unbiased for testing H_o against H_1. It is worth noting that for testing H_o we have tacitly excluded the case that the treatment effect is negative, i.e. $\theta < 0$. It follows that the power of the test (8.7) is less than the level if significance α for $\theta < 0$. Thus if the hypothesis H_o is replaced by H_o': $\theta \leq 0$ that the treatment is no better than the control the test (8.7) of size α is also a level α test of H_o'.

The overall comparison of the Wilcoxon test over its normal theory counterpart the Student-t test suggests that the Wilcoxon test tends to be superior to the t-test in sufficiently non-normal situations. The t-test is more powerful when the distribution F is normal and the loss of efficiency of the Wilcoxon test for normal

F is about five percent. For distributions close to normal (for example a normal distribution with a small proportion of moderate gross errors) there is practically no difference between the powers of the two tests. However if nothing is known about the shape of F and particularly for F with tails heavier than the normal distribution the Wilcoxon test may be considerably more efficient than the t-test (for details see Lehmann(1975)).

To evaluate the constant C extensive tables are available, for example, by Buckle, Kraft and Van Eeden (1969), Milton (1964) and Wilcoxon, Katti and Wilcox (1970) and Lehmann (1975).

In actual practice we very often meet with a difficulty of tie ranks which arises when two more observations are equal. In such a situation it is customary to assign the equal observations their average rank or which is commonly called the mid rank.

Example 8.2. Let $m = 2$, $n = 3$. The observations are 2.2, 2.6, 2.6, 3.1, 3.5. The smallest observation 2.2 has rank 1, the largest observation 3.5 has rank 5, the second largest observation 3.1 has rank 4. We do not know which of the two observations 2.6 has rank 2 and which has rank 3. We assign to each of the observations 2.6 the mid rank $\frac{1}{2}(2+3)=2.5$.

Under $H_0: \theta = 0$ it can be verified that $E(W_{X,Y})=\frac{1}{2}nm$, $\text{Var}(W_{X,Y}) = nm(n+m+1)/12$. Thus from (8.5) under H_0 $E(R)=\frac{1}{2}n(n+m+1)$, $\text{Var}(R)=nm(n+m+1)/12$. For n,m both large (see for example Lehmann (1975))

$$Z = \frac{R - \frac{1}{2}n(n+m+1)}{\sqrt{\frac{1}{12}(nm)(n+m+1)}}$$

is normally distributed with mean 0 and variance 1. In the case of tied observations let R^* be the sum of mid ranks of $Y_1,\ldots,Y_n$. Under H_0 the distribution R^* depends not only on n,m but also on the number of observations tied at each value. Tables of distribution of the Wilcoxon's statistic for tied observations are not practically feasible. So the normal approximation in this case (if applicable) is extremely useful. Under H_o it can be verified that $E(R^*) = \frac{1}{2} n(n+m+1)$. In other words the expected value of the Mann-Whitney statistic remains unaffected for tied observations. But the variance of R^* depends on the number of observations tied at different values. Assume that n+m observation $X_1,\ldots,X_m$; $Y_1,\ldots,Y_n$ take on only p distinct values. Let k_1 of the n+m values be equal to the smallest value, k_2 of the n+m values be equal to the next smallest value,..., k_p to the largest. Then

$$\mathrm{Var}(R^*) = \frac{nm(n+m+1)}{12} - \frac{nm \sum_1^p (k_i^3-k_i)}{12(n+m)(n+m-1)} . \qquad (8.7a)$$

In general the second term in the right hand side of (8.7a) (correction term) is quite small although exceptions are not completely ruled out. If $\max_i k_i/(n+m)$ is bounded away from 1 as $n + m \to \infty$ then under H_o $[R^*-E(R^*)]/\mathrm{Var}(R^*)$ is also normally distributed with mean 0 and variance 1 as $m \to \infty$, $n \to \infty$.

Typically the normal approximation for R^* is not as close as that of R for the same group sizes. The ties tend to decrease the number of values taken on by R and thereby decrease the lumpiness of its distribution. Lehmann (1961) has done a number of comparisons of the exact and the approximate distribution of R with and without continuity correction and has observed that these comparisons are on the whole favourable to the continuity correction.

For testing $H_o: \theta = 0$ against two sided alternatives $H_1: \theta \neq 0$ the Wilcoxon test based on R rejects H_o whenever

$$R \leq C_1 \quad \text{or} \quad R \geq C_2 \quad (C_1 < C_2) \qquad (8.8)$$

where C_1, C_2 are constants such that $P_{H_o}(R \leq C_1) + P_{H_o}(R \geq C_2) = \alpha$ when H_o is true. In cases where two treatments play symmetric role it is natural to choose C_1, C_2 such that under H_o

$$P_{H_o}(R \leq C_1) = P_{H_o}(R \geq C_2) = \frac{\alpha}{2} . \qquad (8.8a)$$

For the fact that the distribution of R is symmetric about $\frac{1}{2}n(n+m+1)$ we can have from (8.8) and (8.8a) that the Wilcoxon test based on R rejects H_o whenever

$$|R-\tfrac{1}{2}n(n+m+1)| \ge C \quad \text{or equivalently} \quad |W_{X,Y}-\tfrac{1}{2}nm| \ge C \qquad (8.8b)$$

where $P_{H_o}\{|W_{X,Y}-\frac{1}{2}mn| \ge C\} = \alpha$. In the presence of ties we replace R by R* in (8.8). Since the distribution of R* is not symmetric we cannot have (8.8b) in terms of R*. But instead we can use (8.8a) to get C_1 and C_2. For large m, n we can use the normal approximation under H_o, as discussed above to determine C_1 and C_2 in both cases.

8.3. ONE-WAY CLASSIFICATION (LOCATION MODEL)

Consider $p(\ge 2)$ independent sets of random variables X_{ij}, $j = 1,\ldots,n_i$, $i = 1,\ldots,p$. Let $X_{i1},\ldots,X_{in_i}$ be a random sample of size n_i from the continuous distributions $F_i(x)$, $i = 1,\ldots,p$ and let

$$F_i(x) = F(x-\theta_i) \text{ for all } x. \qquad (8.9)$$

We first consider the problem of estimating the differences $\Delta_{ij} = \theta_i-\theta_j$. From (8.4)

$$\hat{\Delta}_{ij} = \text{med } (X_{i\alpha}-X_{j\beta}) \qquad (8.10)$$

the median of the $n_i n_j$ differences $X_{i\alpha}-X_{j\beta}$, $\alpha = 1,\ldots,n_i$, $\beta = 1,\ldots,n_j$ is a reasonable estimator of Δ_{ij}. However these estimators are not compatible in the following sense. Suppose that we want to estimate $\theta_2-\theta_1$. When we estimate it directly we get $\hat{\Delta}_{21}$. Since $\theta_2-\theta_1=\theta_3-\theta_1+\theta_2-\theta_3$ we can take estimator of Δ_{21} as the sum of estimators $\hat{\Delta}_{31}$, $\hat{\Delta}_{23}$ of Δ_{31} and Δ_{23} respectively. Obviously $\hat{\Delta}_{21}$, $\hat{\Delta}_{31}+\hat{\Delta}_{23}$ may differ. To avoid this awkwardness Lehmann (1963) proposed to replace (8.10) by

$$\hat{\Delta}_{ij} = \hat{\Delta}_{i.} - \hat{\Delta}_{.j} \qquad (8.11)$$

where $\hat{\Delta}_{k.} = \sum_j \Delta_{kj}/p$ and showed that these adjusted estimators of

Δ_{ij} have essentially the same properties as those of Δ_{ij} in (8.10). If the sample size n_i are not equal Spjøtvoll (1968) suggested the use of

$$\hat{\hat{\Delta}}_{ij} = \hat{\hat{\Delta}}_{i.} - \hat{\hat{\Delta}}_{.j} \qquad (8.12)$$

where

$$\hat{\hat{\Delta}}_{k.} = \sum_j n_j \hat{\Delta}_{kj} \Big/ \sum_j n_j \ .$$

To estimate a contrast $\phi = \sum_1^p \ell_j \theta_j$ subject to the restriction $\sum_j \ell_j = 0$ we first express it in terms of the elementary contrasts Δ_{ij} as $\phi = \sum \sum d_{ij} \Delta_{ij}$ (say). An estimate of ϕ is given by $\hat{\phi} = \sum_i \sum_j d_{ij} (\hat{\Delta}_{i.} - \hat{\Delta}_{j.})$. Lehmann (1963) has shown that $\hat{\phi}$ is uniquely defined (though the d_{ij} are not) and the relative efficiency of $\hat{\phi}$ relative to the normal theory classical estimator $\sum_i \sum_j d_{ij} (X_{i.} - X_{j.})$ is the same as that of $\hat{\Delta}_{ij}$ relative to $X_{i.} - X_{j.}$. In the case of unequal sample sizes it is preferable to use $\hat{\hat{\phi}} = \sum \sum d_{ij} (\hat{\hat{\Delta}}_{i.} - \hat{\hat{\Delta}}_{j.})$. The large sample solutions of the problem of finding simultaneous confidence intervals for all contrasts are discussed by Sen (1966), Puri and Sen (1971), Marascuilo (1971).

Tests of hypotheses. Let $N = \sum_i^p n_i$. Let us rank N observations $X_{ij}, i = 1,\ldots,p,\ j = 1,\ldots,n_i$ is ascending order of magnitude and let R_{ij} denote the rank of X_{ij}. We are interested here in testing the null hypothesis $H_o : \theta_1 = \ldots = \theta_p$. In deciding on an appropriate test we should also need to specify the alternatives against which H_o is being tested. In the analysis of variance model the θ_i represent the responses of p different treatments under consideration and we assume that the treatments mainly affect the response levels. Under this model there is a natural order among the treatments in the sense that one tends to give the lowest response, another gives the next lowest and so on. An indication of the position of a particular treatment in this ordering is provided by its average rank. The average rank for the ith treatment is given by

$$R_{i.} = \sum_j R_{ij}/n_i \, .$$

When H_o is true the $R_{i.}$ will be close to $R_{..} = \sum_i \sum_j R_{ij}/N$. The validity of the hypothesis H_o is obtained by noting how close the averages $R_{i.}$ are to $R_{..}$. The Kruskal-Wallis test based on the $R_{i.}$ is given by rejecting H_o whenever

$$Q = \frac{12}{N(N+1)} \sum_1^p (R_{i.} - \tfrac{1}{2}(N+1))^2$$

$$= \frac{12}{N(N+1)} \sum_1^p \frac{R_i^2}{n_i} - 3\,(N+1) \geq C \tag{8.13}$$

where $R_i = n_i R_{i.}$, and the constant C depends on the size α of the test. For $p = 2$ the statistic Q reduces to the square of the Wilcoxon statistic $W_{X,Y}$. Under H_o

$$P\{R_{ij} = r_{ij},\ j = 1,\ldots,n_i,\ i = 1,\ldots,p\} = \left[\binom{N}{n_1 \ldots n_p}\right]^{-1} \tag{8.14}$$

To determine the value of C corresponding to the level of significance α it is necessary to compute the distribution of Q under H_o which is obtainable from (8.14). Unfortunately except for small values of N the amount of labour required is prohibitive. For $p = 3$, $n_i \leq 5$ and close to 1%, 5% and 10% the values of C have been tabulated by Kruskal and Wallis (1952). Kraft and Van Eeden (1968). Except for very small values of N the distribution of Q can be approximated by a central chi-square distribution with p-1 degrees of freedom. The study of the power function of the Kruskal-Wallis test is quite involved. However Andrew (1954) has shown that the Pitman efficiency of the Kruskal-Wallis test for the model (8.6) relative to its normal theory counterpart, the F test, is the same as the Pitman efficiency of the two sample Wilcoxon test relative to its normal theory counterpart, the t-test.

To deal with cases of tied observations let us suppose that the N observations X_{ij} take on n distinct values. Let k_1 observations are equal to the smallest value, k_2 observations are equal to the second smallest,..., and k_n observations are equal to the largest. Let R^*_{ij} be the midrank of X_{ij} and $R^*_i = \sum_j R^*_{ij}$. The Kruskal and Wallis test of H_o for tied observations is given by rejecting H_o whenever

$$Q^* = \frac{\frac{12}{N(N+1)} \sum_i \frac{R_i^{*2}}{n_i} - 3(N+1)}{1 - \sum_i (k_i^3 - k_i)/(N^3 - N)} \geq C \tag{8.15}$$

where the constant C depends on the size α of the test. Obviously the computation of the distribution of Q^* under H_o is very laborious and tables are impracticable. However except for very small values of N the distribution of Q^* under H_o can be approximated by the central chi-square distribution with p-1 degrees of freedom.

If the hypothesis $H_o: \theta_1 = \ldots = \theta_p$ is rejected by the Kruskal and Wallis test there is a necessity of dividing the treatments into groups, such that the treatments having identical responses belong to the same group and those having different responses belong to different groups; and to find the best treatment. We shall now discuss these below.

Comparison of new treatments with a control. Suppose that the treatments under consideration consists of a standard treatment (control) and (p-1) new treatments which are possible improvements over the standard one. To find the group of new treatments which are better than the control we can apply the Wilcoxon test (Section 8.2) to compare each new treatment with the control. Assume without any loss of generality that the ith new treatment has n_i, i = 2,...,p, observations and the control one has n_1 observations. Thus to apply the Wilcoxon test we combine the n_i observations for the ith new one with n_1 observations for the control and rank them. Let $\bar{R}_i$, i = 2,...,p be the sum of the ranks of n_i observations for the ith new treatment. It may be noted that the $\bar{R}_i$ is different from the R_i as defined in (8.13) which are obtained from the combined ranking of all $N = n_1 + \dots + n_p$ observations. We judge the ith new treatment to be superior to the control of

$$\bar{R}_i \geq C_i, \quad i = 2, \dots, p \tag{8.16}$$

where the constant C_i depends on the level of significance α_i of the test and are determined from the table of the Wilcoxon statistic $W_{X,Y}$. In particular if $n_2 = \dots = n_p$ and if the new treatments play a sufficiently symmetric role, one would then be inclined to choose $\alpha_2 = \dots = \alpha_p = \alpha$ (say), the C_i will then be equal, that is, $C_2 = \dots = C_p = C$ (say), and (8.16) reduces to

$$\bar{R}_i \geq C \tag{8.17}$$

where the constant C is determined such that $P\{R_i \geq C\} = \alpha$ under the hypothesis that the ith new treatment is not superior tc the control. In terms of the shift model (8.9) our aim is to select those new treatments for which $\theta_i - \theta_1 > 0$. The test procedure (8.17) achieves this aim by performing a set of tests of hypotheses $\theta_i - \theta_1 \leq 0$ against the alternatives $\theta_i - \theta_1 > 0$ such that the maximum probability of falsely rejecting at least one of the hypotheses $\theta_i - \theta_1 \leq 0$, $i = 2, \dots, p$ occurs when $\theta_1 = \dots = \theta_p$.

Selection of the best treatment. In comparing p treatments it is natural to face the problem of selecting the best of a number of treatments. The simplest procedure is to select the one with the largest average rank as obtained from the combined ranking of $N = n_1 + \dots + n_p$ observations. In actual practice sometimes it is advantageous to select first a subgroup that is reasonably certain to contain the best treatment. Having narrowed down the choice to this subgroup one would then carry out more extensive experiment to select the best in this subgroup. For selecting the subgroup Gupta and McDonald (1970) suggested the following three rules: Include the ith treatment in the subgroup if

(1) $R_{i.} \geq \max_i R_{i.} - C$ (8.18)

(2) $R_{i.} \geq \dfrac{\max R_{i.}}{C}$ (8.19)

(3) $R_{i.} \geq C$ (8.20)

where C is a constant such that

P{correct selection} $\geq \gamma$ (specified). (8.21)

For the shift model (8.9) with additional assumption that $n_1 = \ldots = n_p = n$ (say) and $\theta_1 < \theta_2 < \ldots < \theta_p$ Gupta and McDonald have shown that for any fixed value of θ_p the probability of correct selection with the rule (8.20) decreases as $\theta_1, \ldots, \theta_{p-1}$ increases towards θ_p and hence the probability of correct selection tends to its minimum value as $\theta_1, \ldots, \theta_{p-1}$ tends to θ_p. The value of C with the rule (8.20) is given by $P\{R_{p.} \geq C\} = \gamma$ whenever $\theta_1 = \ldots = \theta_p$ where $nR_{p.}$ is the sum of ranks of $X_{p1}, \ldots, X_{pn}$ in the joint ranking of the X_{ij} and has the distribution of the Wilcoxon statistic with group sizes $m = n(p-1), n$. The determination of C for the rules (8.18) and (8.19) is not very straightforward and we refer to Gupta and McDonald (1970) for these determinations. Lehmann (1963a), Rizvi and Woodworth (1970) and Ghosh (1973) have also discussed this problem from a different view point.

8.4. EXPERIMENTS WITH TWO OR MORE FACTORS

In such experiments two or more factors each at two or more levels are considered simultaneously. Some of the factors are of real interest to the experimenter while the others are introduced to reduce the variability of the experimental materials and thereby increase the accuracy of the estimators and the statistical test procedures. For the simplicity of presentation we shall consider only the complete block designs in details. However we shall give proper references to incomplete blocks and nonorthogonal design. In deriving suitable nonparametric test statistics we shall consider the method of "n-ranking" and the method of "aligning" only.

Let us now consider the model of a two-way classification where m different treatments are applied to each of n different blocks so that each block contains exactly m treatments and each treatment is replicated exactly n times. Let X_{ij}, $j = 1,\ldots,m$, $i = 1,\ldots,n$ be the random variable associated with the yield of the jth treatment in the ith block. Let X_{ij} be independent and let $F_{ij}(x)=P\{X_{ij} \leq x\}$.

In the parametric case, as we have seen, $F_{ij}(x)=F(x-\mu-\beta_i-\theta_j)$ with $\sum_i \beta_i = 0$, $\sum_j \theta_j = 0$ where the β_i is the ith block effect, the θ_j is the jth treatment effect and μ is the grand mean and also we require the assumption that $F(x)$ is the distribution function of a normal random variable with mean 0 and variance σ^2. The null hypothesis in this case is $\theta_1 = \ldots = \theta_m = 0$ which in this additive model implies that

$$F_{ij}(x) = \ldots = F_{im}(x) = F(x-\mu-\beta_i) .$$

The method of n-ranking assumes neither the additivity nor that the F_{ij} differ only in location. The normality of the F_{ij}, the additivity of the β_i and the homocadesticity of the errors are not required. We shall assume throughout that the F_{ij} are continuous. Under this assumption we formulate the null hypothesis H_o of no treatment effects as

$$H_o\colon\ F_{i1}(x) = \ldots = F_{im}(x) = F_i(x) \quad \text{(say)} \tag{8.22}$$

$i = 1,\ldots,n$. In this formulation the treatment j is stochastically better or worse than the treatment j' if

$$P\{X_{ij} > X_{ij'}\} \geq \tfrac{1}{2} \text{ or } \leq \tfrac{1}{2}$$

for all i with the strict inequality holding for atleast one i. The hypotheses alternative to H_o is H_1: atleast one of the treatments is stochastically better or worse than the rest. In the nonparametric model the specifications of H_o and H_1 become quite simpler if we assume the additivity of the treatment effects in the sense that

$$F_{ij}(x) = F_i(x-\theta_j), \quad j = 1,\ldots,m, \quad i = 1,\ldots,n \tag{8.23}$$

In such cases H_o becomes

$$H_o: \theta_1 \ldots = \theta_m = 0 \tag{8.24}$$

and $H_1: \theta_j \neq 0$ at least for one j. Let R_{ij} be the rank of X_{ij} among $X_{i1}, \ldots, X_{im}$. Since F_{ij} are continuous by assumption ties among observations may be ignored in probability. Let $\underline{R}_i = (R_{i1}, \ldots, R_{im})$. Under H_o the distribution of $\underline{R}_i$ over m! possible realizations is uniform whatever may be the form of F_i provided that they are continuous.

Various tests based on the R_{ij} have been proposed. But the one which is often used is proposed by Friedman (1937). We refer to Puri and Sen (1971) and Lehmann (1975), Kendall (1970) for other tests and relevant results. The Friedman tests based on R_{ij} rejects H_o whenever

$$R = \frac{12}{nm(m+1)} \sum_{j=1}^{m} \left(R_j - \frac{n(m+1)}{2}\right)^2$$

$$= \frac{12}{nm(m+1)} \sum_{1}^{m} R_j^2 - 3n(m+1) \geq C \tag{8.25}$$

where $R_j = \sum_{i=1}^{n} R_{ij}$, the sum of the ranks for the treatment j and the constant C depends on the level of significance α of the test. Tables of the distribution of Q under H_o is given by Friedman (1937), Owen (1962), Kraft and Van Eeden (1968) and Kendall (1970). For n and m large Q is distributed as χ^2_{m-1} (chi-square with m-1 degrees of freedom) whenever H_o holds.

For tied observations in the same block we assign the tied observations the midranks (the average of the ranks in the same block with which the observations are tied). Ties among observations in different blocks cause no difficulty since the observations in the same block are being ranked to obtain the test statistic Q. Suppose the m observations in the ith block take t_i distinct values with d_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, t_i$, of the observations being equal to the jth largest value among the t_i distinct values in the ith block. Let R^*_{ij} be the midrank of the observation X_{ij} and let $R^*_j = \sum_{i=1}^{n} R^*_{ij}$. The hypothesis H_o with the tied observations is rejected whenever

$$Q^* = \left[\frac{12}{nm(m+1)} \sum_1^m R_j^{*2} - 3n(m+1)\right] \Big/ \left[1 - \sum_i \sum_j (d_{ij}^3 - d_{ij})/nm(m^2-1)\right] \geq C \qquad (8.26)$$

where the constant C depends on the level of significance α of the test. The distribution of Q^* under H_o is again χ^2_{m-1} when n and m are large. Typically this chi-square approximation for both Q and Q^* is quite reasonable whenever $mn \geq 30$. The Friedman test for $m = 2$ reduces to the sign test and the Wilcoxon's sign rank test is more effective for this. However it has been observed that the Friedman test is relatively less sensitive for low values of m and it is mainly due to the fact that separate ranking in each block provides comparisons only the responses within each block. Because of the variations between blocks, of which some may give consistently high and others consistently low responses, a direct comparison of responses in different blocks is not, in general, very meaningful. This difficulty can be eliminated by subtracting from the observations in each block some estimate of the location in each block such as the average of the observations of the block or the median of these observations. This technique is commonly called the method of aligning. Contrary to the method of n-ranking the method of aligning assumes the additivity of the block effects and the homoscedasticity of the errors of different blocks. In other words to use the technique of aligning we assume that

$$X_{ij} = \mu + \alpha_i + \theta_j + \ell_{ij}, \quad i=1,\ldots,n, \; j=1,\ldots,m \qquad (8.27)$$

where μ is the grand mean, $\alpha_1,\ldots,\alpha_n$ are the block effects satisfying $\Sigma\alpha_i = 0$ (may be random or fixed), $\theta_1,\ldots,\theta_m$ are the treatment effects satisfying $\Sigma\theta_j = 0$ (fixed only) and the ℓ_{ij} are the error components. It is assumed that

$$\underline{\ell}_i = (\ell_{i1},\ldots,\ell_{im})', \quad i = 1,\ldots,n$$

are independent random vectors having continuous distribution function $F_i(x_1,\ldots,x_m)$, $i = 1,\ldots,n$ such that each F_i is symmetric in each of its m arguments. Obviously this includes as a special case the classical case of independent and identically distributed error components ℓ_{ij}.

The technique of aligning consists of eliminating the block effects α_i by taking

$$Y_{ij} = X_{ij} - \tilde{X}_i \tag{8.28}$$

where $\tilde{X}_i$ is a suitable estimator of the location of the ith block. We shall consider here the case $\tilde{X}_i = X_{i.}$, the average of the ith block. Let us now arrange $Y_{ij} = X_{ij} - X_{i.}$, $i = 1,\ldots,n$, $j = 1,\ldots,m$ in order of magnitudes and let $\hat{R}_{ij}$ be the rank (or mid-rank in the case of tied Y_{ij}). The $\hat{R}_{ij}$ are called the aligned rank of the Y_{ij}. Under $H_o{:}\,\theta_1=\ldots=\theta_m=0$, that is, under the hypothesis of no treatment effect regardless of which position in a block were assigned to which treatment, the m! possible arrangements in any given block will simply result in m! permutation of the aligned ranks in this block. The total number of equally likely permutations within the n sets of aligned ranks in all blocks is $(m!)^n$. Therefore the distribution of $\hat{R}_{ij}$ under H_o is given by

$$P\{\hat{R}_{ij} = \hat{r}_{ij},\ i = 1,\ldots,n,\ j = 1,\ldots,m\} = (m!)^{-n}.$$

Let $\hat{R}_j = \sum_i \hat{R}_{ij}$, $R_{j.} = \frac{1}{n}\hat{R}_j$. Under H_o $\sum \hat{R}_{j.} = \frac{1}{2}m(mn+1)$. A test based on $\hat{R}_{ij}$ of H_o which is analogous to the Friedman Q test rejects H_o whenever

$$\hat{Q} = \frac{m-1}{\sum_i\sum_j(\hat{R}_{ij}-\hat{R}_{i.})^2}\left[\sum_{j=1}^{m}\left(\hat{R}_j-\tfrac{1}{2}n(n\ m+1)\right)^2\right]$$

$$= \frac{m-1}{\sum_i\sum_j(\hat{R}_{ij}-\hat{R}_{i.})^2}\left[\sum_{j=1}^{m}\hat{R}_j^2-\tfrac{1}{4}mn^2(m\ n+1)^2\right]$$

$$\geq C$$

where $\bar{R}_{i.} = \frac{1}{m}\sum_j \hat{R}_{ij}$ and the constant C depends on the level of significance α of the test. Since $\sum_i\sum_j \hat{R}_{ij}^2$ is the sum of squares of all ranks or mid-ranks, it is a constant. Similarly, as the set of m ranks in a block is fixed (only the order being random), the $m\hat{R}_{i.}$ being the sum of the ranks of the ith block are also constants. Hence

$$\sum_i\sum_j (\hat{R}_{ij}-\hat{R}_{i.})^2 = \sum_i\sum_j \hat{R}_{ij}^2 - \frac{1}{m}\sum_i (m\,\hat{R}_{i.})^2$$

is a constant.

For small values of m, n the distribution of $\hat{Q}$ under H_o is very complicated. However for large values of m, n $\hat{Q}$ is distributed as χ^2_{m-1}. For computational purposes, the test based on $\hat{Q}$ is more laborious than the test based on Q. To see if this extra effort is warranted it is necessary to know by how much it will increase the efficiency of this procedure. Under the assumption of additivity and under the assumption of normality of $F_{ij}(x)$ the optimum test of $H_o:\theta_1=\ldots=\theta_m=0$ rejects H_o whenever

$$U = \frac{\sum_j n(X_{.j}-X_{..})^2/(m-1)}{\sum_i\sum_j (X_{ij}-X_{.j})^2/(m-1)(n-1)} \geq C \tag{8.30}$$

where the constant C depends on the size α of the test and under H_o the statistic U has central F-distribution with (m-1), (n-1)(m-1) degrees of freedom. We are interested in the Friedman test based on Q and $\hat{Q}$ mainly because we are not sure about the normality assumption of the distribution function F. Since the distribution of Q (or $\hat{Q}$) under H_o is independent of the form of the distribution function F, the level of significance α of the Friedman test based on Q (or $\hat{Q}$) is thus independent of the form of the distribution function F. Obviously this is not true for the test based on U where the level of significance does depend on the form of distribution function F. However if $n \to \infty$ the distribution of U under H_o tends to χ^2_{m-1} irrespective of the form of F provided the variance of the distribution F is finite. To answer the question raised above in connection with the test based on $\hat{Q}$ one should necessarily compare the Friedman test with the normal theory F-test based on U in terms of their Pitman efficiency which is defined as the limiting value as $n \to \infty$ of the ratio of the number of blocks required by the two tests of the same size α to achieve the same power against the same alternatives. Van Elteren and Noether (1959) showed that the Pitman efficiency of the Friedman test relative to the F-test is 0.637 for m = 2, 0.716 for m = 3, 0.764 for m = 4, 0.769 for m = 5, 0.868 for m = 10, 0.895 for m = 15 and 0.955 for m → ∞. Hence the need of improvement of the Friedman test through alignment is greatest if m is small and is less important when m is large. Mehra and Sarangi (1967) have shown that the Pitman efficiency of the test based on $\hat{Q}$ relative to the Friedman test based on Q decreases as m increases and that it is equal to 1.5 for m = 2, 1.355 for m = 3, 1.263 for m = 4, 1.210 for m = 5 and 1 for m → ∞. In this context it is worth noting that the test based on Q is equivalent to the Wilcoxon test for the two-sample problem described earlier in this chapter.

So far we have considered in this section complete two-factor layout, that is, the size of the blocks is equal to the number of treatments being compared. However it is frequently necessary to have blocks size less than m such as balanced incomplete block designs or partially incomplete block designs, etc. Conversely it may also happen that the size of the blocks is greater than m. The Friedman test for the absence of treatment effects is generalized to cases where the block size is different than m and in particular by Durbin (1951), Bernard and van Eltern (1953), and Sen (1968a, 1968b, 1977) to the balanced incomplete block designs. Mehra and Sarangi (1967), and Sarangi and Mehra (1969) have generalized the test based on $\hat{Q}$ to the case where the number of observation n_{ij} on the jth treatment in the ith block is not necessarily equal to 1 (nonorthogonal data).

Consider a balanced incomplete block design with m treatments arranged in b blocks each containing k treatments $(k \leq m)$ and each treatment is replicated $r(\leq b)$ times in such a way that no treatment occurs more than once in any block and each pair of treatments occurs together exactly in λ blocks. If the jth treatment occurs in the ith block we denote by R_{ij} the rank of its corresponding response among the k responses in the ith block. Otherwise we let $R_{ij} = 0$. Let $R_j = \sum_i R_{ij}$, $j = 1,\ldots,m$. For testing the null hypothesis H_o of no treatment effects the test analogous to the Friedman test rejects H_o whenever

$$Q = \frac{12}{m\lambda(k+1)} \sum_{j=1}^{m} [R_j - \tfrac{1}{2}r(k+1)]^2 \geq C$$

where the constant C depends on the level of significance α of the test and under H_o Q is distributed as χ^2_{m-1} when both n,m are large.

EXERCISE 8

1. Show that if the distribution of $\underset{\sim}{X}_1,\ldots,X_n$ is symmetric about θ then the distribution of $\bar{\theta}, \theta$ and $\hat{\theta}$ of section 8.1 are also symmetric about θ.

2. Show that in the two-sample shift model

(a) $\hat{\theta}-\theta$ is independent of θ when F is continuous.

(b) $\hat{\theta}$ is symmetrically distributed about θ if either the distribution of F is symmetric about θ or if m = n.

3. Show that under H_o: θ= 0 $E(W_{X,Y}) = \tfrac{1}{2}mn$, $Var(W_{X,Y}) = \frac{1}{12} nm(n+m+1)$.

4. Source: Davies, O.L. (1954), Design and analysis of industrial experiments, Oliver and Boyd, London.

The following data were collected from a wear-testing experiment with 4 materials M_1, M_2, M_3, M_4. A test piece of each material was taken from each of the 3 positions P_1, P_2, P_3 of a testing machine. The loss in weight due to wear was measured on each piece of material in milligrams.

Table 8.1. Loss in weight due to wear

	M_1	M_2	M_3	M_4
P_1	241	195	235	234
P_2	270	241	273	236
P_3	274	218	230	227

Analyse the data to test if there are significant differences due to materials.

5. The following observations were made from an experiment conducted in West Bengal, India with 4 different varieties of Wheat on the average number of grain per 10 ears of 14 randomly selected plants for each variety.

Table 8.2. Average Number of Grains per 10 ears

Plant number	Variety 1	Variety 2	Variety 3	Variety 4
1	261	351	322	323
2	320	424	380	310
3	287	370	260	310
4	308	328	310	395
5	311	380	278	376
6	281	374	306	401
7	254	357	307	412
8	346	282	346	389
9	290	404	337	368
10	287	412	285	356
11	320	390	272	328
12	307	400	276	360
13	328	406	280	405
14	339	300	290	412

Analyse the data to test the null hypothesis of no variety effect.

6. The following data on the average length of ear (in cm) for each of 10 randomly selected plants were recorded from an experiment with 3 varieties of barley crop.

Table 8.3. Average length of ears (in cm)

Plants number	Variety 1	Variety 2	Variety 3
1	8.97	10.19	9.65
2	9.63	9.47	9.76
3	9.59	10.75	10.52
4	9.58	10.48	9.78
5	9.02	10.21	9.30
6	14.40	10.79	9.80
7	10.58	11.06	10.30
8	9.32	11.14	9.32
9	9.50	11.43	9.68
10	10.40	11.12	9.70

Analyse the data to test the null hypothesis of no variety effect.

7. Source: Damaser, Shore and Orne (1963). Psychological effects during hyphonically required emotions, Psychosomatic medicine, 25, pp. 334-343 (also reported in Lehmann (1975)).

In a study of hypnosis the emotions of fear, happiness, depression and calmness were requested in a random order from each of 8 subjects during hypnosis. Table 8.4 gives the resulting measurements of skin potential adjusted to initial level in millivolts.

Table 8.4. Skin potential in millivolts

Emotions	Subjects 1	2	3	4	5	6	7
Fear	23.1	57.6	10.5	23.6	11.9	54.6	21.0
Happiness	22.7	53.2	9.7	19.6	13.8	47.1	13.6
Depression	22.5	53.7	10.8	21.1	13.7	39.2	13.7
Calmness	22.6	53.1	8.3	21.6	13.3	37.0	14.8

Analyse the data to test the hypothesis of no emotion effect.

8. Source: C.W. Loel (1952) Industrial statistics, Business Books Ltd., 180 Fleet Street London EC4.G.B.

Chemical reaction take place in solutions and when one of the reactants is also a good solvent for the other reactants it is customary to run the reaction in presence of a large excess of solvent reactants. At the end of the reaction the excess of the solvent reactant has to be recovered for reuse and often it is found to be contaminated with some impurity formed in the reaction. If this impurity content increases on repeated recovery, it may inhibit the reaction and it is necessary to clean the solvent by some means. One way in which an estimate may be made of whether or not the solvent needs cleansing is to check the reaction batch time cycles. Table 8.5 exhibits the data collected from an experiment in which five reactors R_1, R_2, R_3, R_4, R_5 and runs

Table 8.5. Order of starting batches

Block	1	2	3	4	5
1.	(R_5) 18.4	(R_3) 18.0	(R_2) 18.4	(R_1) 18.2	(R_4) 18.0
2.	(R_4) 18.4	(R_1) 18.6	(R_5) 19.0	(R_2) 19.0	(R_3) 18.6
3.	(R_2) 19.5	(R_4) 18.9	(R_3) 19.6	(R_5) 20.0	(R_1) 19.3

were blocked so that Block 1 deals with runs made with fresh solvent, Block 2 deals with runs made with some fresh and some recovered solvent and Block 3 deals with runs with some fresh and some solvent which has been through the process twice. Batches were assigned to the various reactors in random order. Analyse the data to test the hypothesis of no block effects.

REFERENCES

Andrews, F.C. (1954). Asymptotic behaviour of rank sum tests for analysis of variance, Ann. Math. Statist., 25, pp.724-736.

Benard, A. and Van Elteren, P. (1953). A generalization of the method of m-rankings, Indagationes, Maths., 15, pp.358-369.

Buckle, N., Kraft, C.H. and Van Eeden, C. (1969). Tables prolongees de la distribution de Wilcoxon-Mann-Whitney. Presse de l'Universite de Montreal, Montreal.

Durbin, J. (1951). Incomplete blocks in ranking experiments, Brit. J. Psychol. 4, pp.85-90.

Van Eltern, P. and Noether, G.E. (1959). The asymptotic efficiency of X^2-test for a balanced incomplete block design, Biometrika 46, pp.475-477.

Friedman, M. (1937). The use of ranks to avoid assumption of normality implicit in the analysis of variance, J.A.S.A. 32, pp. 675-701.

Ghosh, M. (1973). Nonparametric selection procedures for symmetric location parameter populations, Ann.Statist. 1, pp.773-779.

Gupta, Shanti and McDonald, G.C. (1970). On some classes of selection procedures based on ranks, Nonparametric Techniques in statistical inference, M.L. Puri (Edited). Cambridge University Press, London, pp.491-514.

Hodges, J.L. and Lehmann, E. (1963). Estimates of location based on ranks, Ann. Math. Statist., 34, pp.598-611.

Kraft, C.H. and Van Eeden, C. (1968). A nonparametric introduction to statistics, Macmillan, N.Y.

Kruskal, W.H. and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis, J.A.S.A. 47, pp.583-612.

Lehmann, E. (1963). Robust estimation in analysis of variance, Ann. Math. Statist. 34, pp.957-966.

Lehmann, E. (1963a). A class of selection procedures based on ranks, Math. Ann., 150, pp.268-275.

Lehmann, E. (1975). Nonparametric statistical methods based on ranks, Holden-Day, San Francisco.

Lehmann, Shirley, Young (1961). Exact and approximate distributions for the Wilcoxon statistic with ties, J.A.S.A., 56, pp.293-298.

Marascuilo, L.A. (1966). Large sample multiple comparisons, Psychol. Bull. 64, pp.280-290.

Mehra, K.L. and Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments, Ann. Math. Statist. 38, pp.90-107.

Milton, R.C. (1964). An extended table of critical values for the Mann-Whitney (Wilcoxon) two-sample statistic, J.A.S.A. 59, pp.925-934.

Owen, D.B. (1962). Handbook of statistical tables, Addison Wesley, Reading, Mass.

Puri, M.L. and Sen, P.K. (1971). Nonparametric methods in multivariate analysis, Wiley, N.Y.

Rizvi, M.H. and Woodworth, G.G. (1970). On selection procedures based on ranks, Ann. Math. Statist. 40, pp.1941-1951.

Sarangi, J. and Mehra, K.L. (1969). Some further results on Hodges-Lehmann conditional rank tests. Bull. Calcutta Statist. Assoc. 18, pp.25-41.

Sen, P.K. (1963). On the estimation of relative potency in dilution (-direct) arrays by distribution free methods, Biometrics, 19, pp.532-552.

called positive definite if $\underline{x}'A\underline{x} > 0$ for all $\underline{x} \neq \underline{0}$ and is

Sen, P.K. (1966). On nonparametric simultaneous confidence regions and tests for the one-criterion analysis of variance problem. Ann. Inst. Statist. Math. 18, pp. 319-336.

Sen, P.K. (1968a). On a class of 'aligned' rank order tests in a two-way layouts, Ann. Math. Statist. 39, pp.1115-1124.

Sen, P.K. (1968b). Robustness of some nonparametric procedures in linear models. Ann.Math.Statist. 39, pp.1913-1933.

Sen, P.K. (1971). Asymptotic efficiency of a class of aligned rank order tests for multiresponse experiments in some incomplete block designs. Ann. Math. Statist. 42, pp.1104-1112.

Wilcoxon, F., Katti, S.K. and Wilcox, R.A. (1970). Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. Selected tables in Mathematical statistics. (Harter & Owen eds.) Markham Publishing Co., Chicago, vol.1, pp.171-259.

AUTHOR INDEX

Allan, F.E., 201,231
Anderson, T.W., 346,354
Andrews, F.C., 409

Bartlett, M.S., 358
Bernard, A., 418
Birkoff, G., 28
Blackwell, D., 61
Box, G.E.P., 352
Bose, R.C., 96,171,227,247,273,275,290, 291
Buckle, N., 405

Chakrabarti, M.C., 171,245
Clatworthy, W.H., 275
Cleroux, R., 289
Cochran, W.G., 200,224,243,270
Consul, P.C., 351
Cornish, E.A., 252,272
Cox, G.M., 200,270

Das, M.N., 227,248,337
Das Gupta, S., 354
DeLury, D.B., 232
Dembowski, P., 275
Duncan, D.B., 133,135,171
Durbin, J., 418

Erhenfeld, S., 190

Federer, W.T., 133,273,304,331,333,334
Ferguson, G.A., 171
Fisher, R.A., 197,200,225,226,227,247
Foster, R.G., 353
Friedman, M., 414

Gabriel, K.R., 353
Gnanadeshikhan, R., 353
Ghosh, J.K., 73
Ghosh, M., 412
Giri, N., 7,13,15,20,34,47,77,79,82,118, 191,227,256,282,290,291,348,354
Graybill, F., 256,264
Gupta, A.K., 351
Gupta, S., 411,412
Guttman, I., 295,296

Hall, W.J., 73
Hannan, E.J., 381
Harrington, H.B., 243
Harshbarger, B., 333
Hartley, H.D., 270
Heck, D.L., 353
Hodges, J.L., 401,402
Hotelling, H., 212,219
Huges, J.B., 353
Hunter, J.S., 295,296
Hsu, P.L., 100,171,342,346

James, A.T., 381
Jayachandran, K., 353
John, P.W.M., 296,334
John, S., 354
Jouris, G.M., 381

Katti, S.K., 405
Keuls, M., 134
Kempthorne, O., 171,304,331,334
Khatri, C.G., 381
Kiefer, J., 77,78,79,188,190, 191,248,255,271,348
Kirkman,T., 247
Kishen, K., 290
Kraft, C.H., 405,409,414
Kruskal, W.H., 409
Kshirsagar, A.M., 255

Lawley, D.N., 346
Lee, Y.S., 352
Lehmann, E., 138,172,212,324,380, 399,401,402,405,406,407,408, 412,414
Lehmann, E.L., 67,77,78,79
Linnik, J.V., 77,348

Ma, R.H., 243
Mangess, T.A., 104
McDonald, G.C., 411,412
MacDuffee, C., 19,27
McGuire, M., 104,105
McLane, S., 27
MacNeish, H.F. , 227
Marascuilo, L.A. ,408
Markus, M., 27
Mann, H.B., 227
Marrow, D.J., 353
Mathai, A.M., 352
Mehra, K.L., 418
Mikhail, W.F., 354
Miller, R.G., 172
Milton, R.C., 405
Mine, H., 27
Mitra, S.K., 268
Mote, V.L., 255
Mudolkar, G.S., 354

Nair, K.R., 232,256,273,291,333,334
Nair, U.S., 351
Nandi, H.K., 354
Narain, R.D., 354
Newman, D., 134
Neyman, J., 190,205
Noether, G.E., 418

Owen, D.B., 414

Parkar, E.T., 227
Pearson, E.S., 190
Perlis, S., 28
Pillai, K.C.S., 346,351,353,381
Pliss, V.A., 77,348
Puri, M.L., 399,408,414

Rao, C.R., 28,100,253,256,264,269, 352,380
Rao, Raghav V., 227,245,248,275, 304,334
Rao, S.V.S.P., 248
Rathie, P.N., 352
Rees, D.D., 353
Rizvi, M.H., 412
Roy, J., 255
Roy, S.N., 28,353,354,380

Salaevskii, O.V., 77,348
Sarangi, J., 418
Schatzoff, M., 381
Scheffe, H., 133,212,219
Schimamoto, T., 273,275
Schwartz, R., 78,348,353,354
Sen, P.K., 399,401,402,408,414,418
Seshadri, V., 256,264
Shah, K.R., 256
Shrikhande, S.S., 227,270,275,290
Sinha, B.K., 354
Smith, H., 353
Sprott, D.A., 264,282
Stein, C., 77,78,79,348
Student, 134

Taylor, W.B., 270
Tocher, K.D., 248
Tukey, J.W., 133,172

Van Eeden, C., 405,409,414
Van Elteren, P., 418
Vartak, M.N., 245

Wald, A., 190
Wallis, W.A., 409
Weeks, D., 256,264
Welch, B.L., 242
Wijsman, R.A., 73
Wilcox, R.A., 405
Wilcoxon, F., 405
Wilks, S.S., 48,295,296,342,346, 352,358
Wishart, J., 192,201,231
Wolfowitz, J., 188
Woodworth, G.G., 412

Yates, F., 226,227,231,232,247, 256,261,264,331,333,342
Youden, W.G., 290
Young, D.L., 381

SUBJECT INDEX

Admissibility, 78
Almost invariance, 66
Analysis of covariance, 174,179, 180,205,272
A-optimality, 255

Balanced designs, 245
Balanced incomplete block designs, 244,246

Characteristic roots, 5
Characteristic vectors, 5
Cochran's theorem, 42
Cofactor, 3
Complete block designs, 196
Completely randomized block designs, 197
Complete matrix, 19
Complex multivariate normal distribution, 380
Confidence interval, 118
Confounding, 310
Connected designs, 245
Contrasts, 244
Crossed designs, 325

Direct product, 24
Distributions
  noncentral beta, 32
  noncentral chisquare, 32
  noncentral F, 32
  noncentral-t, 32
  quadratic forms, 38
Duncan's multiple range, 137

Echelon form, 14
Efficiency of designs, 187
Efficiency of Latin square designs, 242
Estimable function, 96

F-test, 231
Factorial experiments, 304
Fisher-Cochran theorem, 40
Fixed effect model, 257

Gauss-Markoff theorem, 257
General factorial experiments, 307
Generalized inverse, 10
Generalized variance, 45
General linear hypothesis, 87,93
  invariance, 89
General linear regression model, 183
Greco-Latin squares, 226

Groups
  Abelian, 22
  affine, 22
  full linear, 22
  permutation, 22
  subgroup, 23
  translation, 22

Hermit canonical form, 14,20
Homomorphism, 24
Hotelling's T^2, 51,78,347,348
Hunt-Stein theorem, 71,78

Idempotent matrix, 10
Incomplete block designs, 196
Intrablock analysis
  BIB designs, 248
  PBIB designs, 278
Invariance, 54,73
Invariant function, 61
Inverse matrix, 3

Jacobian, 20

Kruskal-Wallis test, 409

Latin square, 226
Lattice design, 332
Lawley-Hotelling trace test, 353
Least squares
  canonical form, 91
  estimators, 97
  estimator (correlated observations), 103
Likelihood ratio test, 121,354

Mann-Whitney statistic, 403
Matrix
  diagonal, 2
  nonsingular, 2
  orthogonal, 2
  positive definite, 4
  triangular, 2
Maximal invariant, 61
Maximum likelihood, 113
Minor, 3
Missing observations, 201,231,252
Mixed effect model, 325
Multinomial distribution, 53,380
Multiple range test, 133
Multivariate analysis of covariance, 379
Multivariate general linear hypothesis,339
  applications, 361

Multivariate classifications
  one-way, 366
  two-way, 368,372
Multivariate randomized block
  design, 376

Nested designs, 314,322,325
Neyman and Pearson's lemma,
  64
Neyman and Pearson's type
  C test, 190
Normal subgroup, 23

One-way classification, 128
  location model, 407
Optimality criteria,
  A, 189
  D, 189
  E, 189
  L, 189
  M, 189
Optimality of designs, 187
Optimum tests, 76

Partially balanced association
  scheme, 273
Partitioned matrices, 17
Permutation test, 211
Pillai's test criterion, 366

Quadratic form, 4
Quotient group, 23

Randomized block design, 199,231
Randomization model, 205,234
Randomization test, 267,289
Range test, 133
Recovery of intrablock information,
  282
Rectangular lattice designs, 333
Regression model, 121
Row balance (PBIB), 289
Roy's test, 353

S-method, the, 133
Splitplot design, 304
Student-Newman-Keul's test, 134
Sufficiency, 73

Tests of equality of variances,138
Three-way classification, 158
Transformation group, 25
Transformation set, 226
Tukey's T-method, 133
Two sample location parameter, 402
Two-way classification, 148,154

Unbiasedness, 73

Wilcoxon's signed rank test statistic,
  401,404
Wishart distribution, 45

ANALYSIS OF VARIANCE

N. Giri, University of Montreal, Montreal, Quebec H37 3J7

This book deals with an up-to-date presentation of both theoretical and applied aspects of analysis of variance in univariate and multivariate cases. Various designs are explained with live data from different applied fields. Each chapter contains numerous related problems, complete references and worked out examples. The exercises of chapters 6 and 7 on multivariate analysis of variance also contain analogous result for the complex multivariate normal populations.

Contents : Some mathematical concepts. Basic statistical concepts. Univariate general linear hypothesis and analysis of variance. Analysis of covariance optimality of designs. Experimental designs and their analysis. Analysis of variance II, factorial experiments. Multivariate general linear hypothesis with applications. Nonparametric analysis of variance. Author index. Subject index.

1986 429 pages **Rs. 36.00**

SOUTH ASIAN PUBLISHERS PVT. LTD.
36 Netaji Subhash Marg, Daryaganj, New Delhi 110002